MAX DELBRÜCK CENTER BIMSB

Berlin Institute for Medical Systems Biology

Discovery for tomorrow's medicine

SEQUENCING PLANET EARTH

18TH BERLIN SUMMER MEETING

19—20 JUNE 2025

Computational & Experimental Molecular Biology Meet

SEQUENCING PLANET EARTH

18TH BERLIN SUMMER MEETING
19—20 JUNE 2025
Computational & Experimental Molecular Biology Meet

MDC-BIMSB
Hannoversche Str. 28, 10115 Berlin

Organization & Contact
Verena Maier
BIMSB.events@mdc-berlin.de
www.mdc-berlin.de/de/BSM2025

W-Lan: mdc_open

MAX DELBRÜCK CENTER BIMSB
Berlin Institute for Medical Systems Biology
Discovery for tomorrow's medicine

Supported by

Novogene

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Dear Friends, Colleagues & Guests,

We are excited to welcome you to the 2025 Berlin Summer Meeting!

Now in its 18th year, this annual international gathering of scientists working at the intersection of computation and experiment in molecular biology has become a vibrant forum for cross-disciplinary exchange, uniting early-career researchers and established experts to share insights and spark collaborations.

This year's theme "Sequencing Planet Earth" captures both the ambition and the urgency of our work. Recent breakthroughs in sequencing technologies are dramatically increasing both throughput and precision. Concurrently, leaps in data science are enabling us to connect, compare, interrogate, and interpret biological data on an unprecedented scale. From understanding biodiversity to combating global health threats, the integration of these tools is opening new pathways for discovery that impact communities and ecosystems across the globe.

Over the next two days, we will hear from researchers who are expanding the boundaries of what's possible – from single-cell transcriptomics to planetary-scale genomics & environmental surveillance, from novel computational models to real-time biological and clinical insights. We are particularly excited to showcase work that not only deepens our scientific

understanding but also addresses shared global challenges through inclusive, collaborative approaches.

Whether you are presenting your latest findings, engaging in poster sessions, or joining conversations over coffee, we encourage you to connect, question, and contribute.

Thank you for joining us — we look forward to a dynamic and productive meeting!

Sincerely,
The Berlin Summer Meeting 2025 Scientific Organizing Committee:

Sofia Forslund Charité – Universitätsmedizin Berlin/Max Delbrück Center
Nick Gilbert University of Edinburgh, MRC Human Genetics Unit
António Jacinto NOVA Institute for Medical Systems Biology (NIMSB)
Beate Kampmann Charité Center for Global Health
Markus Landthaler MDC-BIMSB
Helder Nakaya University of São Paulo
Nikolaus Rajewsky MDC-BIMSB
Juliana Roscito MDC-BIMSB
Ashley Sanders MDC-BIMSB

The visuals for the 18th Berlin Summer Meeting:
To visualize the theme of *Sequencing Planet Earth*, we built a physical model composed of layered sheets of printed acrylic glass. Shaped into a globe, the structure symbolizes the complexity of life, disease, technology, and data. The idea also draws from early X-ray crystallography techniques from the mid-20th-century that mapped molecular structures on stacked glass plates.

SLANG Design Studio
www.slanginternational.org

THURSDAY, JUNE 19, 2025

8:30 am • 9:00 am	**Registration & coffee**
9:00 am • 9:15 am	**Welcome address by Holger Gerhardt,** Deputy Scientific Director of the Max Delbrück Center
	SESSION 1 Sequencing at Scale – Tools to Decode Life's Molecular Code Chair: **Juliana Roscito**
9:15 am • 10:00 am	**Ulf Landegren,** Uppsala University, Sweden *Tools to analyze very many molecules – and even very few*
10:00 am • 10:15 am	**Daniel Léon-Periñán,** MDC-BIMSB, Berlin, Germany *Malva: Real-time Sequence Search across Billions of Cells*
10:15 am • 11:00 am	**Camille Marchet,** CNRS, Université de Lille, France *Scaling genomic reuse: hypothesis and algorithms for k-mer collections*
11:00 am • 11:30 am	**Coffee break**
	SESSION 2 Expanding the Map – Inclusive Transcriptomics Across Species and Populations Chair: **Nick Gilbert**
11:30 am • 12:15 pm	**Vinicius Maracaja-Coutinho,** University of Chile, Santiago, Chile Addressing inclusion in transcriptomic studies: from animal models to human single-cell biology
12:15 pm • 12:30 pm	**Can Ergen,** UC Berkeley, USA ResolVI – addressing noise and bias in spatial transcriptomics
12:30 pm • 12:45 pm	**Tilman Hoffbauer,** Technical University of Munich, Germany DNA language model based modelling of transcription start sites in twelve yeast species
12:45 pm • 1:00 pm	**Olga Bakina,** Charité – Universitätsmedizin Berlin, Germany RNA 4SU metabolic labeling in human retinal organoids under dissociation, glucose deprivation and light exposure
1:00 pm • 1:10 pm	**Group photo**
1:10 pm • 3:00 pm	**Lunch & poster session**

THURSDAY, JUNE 19, 2025

SESSION 3
The RNA World – Mapping Viromes and Health Across Urban and Global Landscapes
Chair: **Markus Landthaler**

3:00 pm • 3:45 pm	**Helder Nakaya,** University of São Paulo, Brazil AI for Health: From Precision Diagnostics to Pandemic Preparedness
3:45 pm • 4:00 pm	**Alexander Lucaci,** Weill Cornell Medicine, New York City, USA Diversity and Distinctive Traits of the Global RNA Virome in Urban Environments
4:00 pm • 4:15 pm	**Udo Gieraths,** Charité – Universitätsmedizin Berlin, Germany A century-old museum sample reveals a bandavirus with modern day presence in northern European bats
4:15 pm • 5:00 pm	**Marion Koopmans,** Erasmus Medical Center, Rotterdam, the Netherlands TBA

FRIDAY, JUNE 20, 2025

8:30 am • 9:00 am	**Registration & coffee**
	SESSION 4 Systems Medicine for a Connected World – Modeling Immune Responses and Disease Outcomes Chair: **Antonio Jacinto**
9:00 am • 9:45 am	**Beate Kampmann,** Charité Center for Global Health, Berlin, Germany *A systems-biology approach to unravel the ontogeny of the immune system in newborns and the impact of vaccination*
9:45 am • 10:00 am	**Natalia von Staa Mansur,** University of São Paulo, Brazil *Graph Neural Networks in Precision Medicine: Predicting Clinical Outcomes of Infectious Diseases*
10:00 am • 10:45 am	**Grace Androga & Brenda Kwambana Adams,** Malawi Liverpool Wellcome Programme, Blantyre, Malawi *Genomics without borders to inform vaccine deployment against outbreak prone bacteria in Africa*
10:45 • 11:15	**Coffee break**
	SESSION 5: Building Genomic Infrastructure for Planet Earth Chair: **Ashley Sanders**
11:15 am • 12:00 pm	**Alfred Ngwa**, MRC Unit, The Gambia, London School of Hygiene & Tropical Medicine, UK *Evolutionary dynamics and malaria transmission in Africa*
12:00 pm • 12:15 pm	**Jonathan Wood,** Wellcome Sanger Institute, Hinxton UK *Tree of Life – generating reference genomes at scale*
12:15 pm • 1:00 pm	**Tobias Marschall,** Heinrich Heine University, Düsseldorf, Germany *Pangenome based analysis of structural variation*
1:00 pm • 2:00 pm	**Lunch**

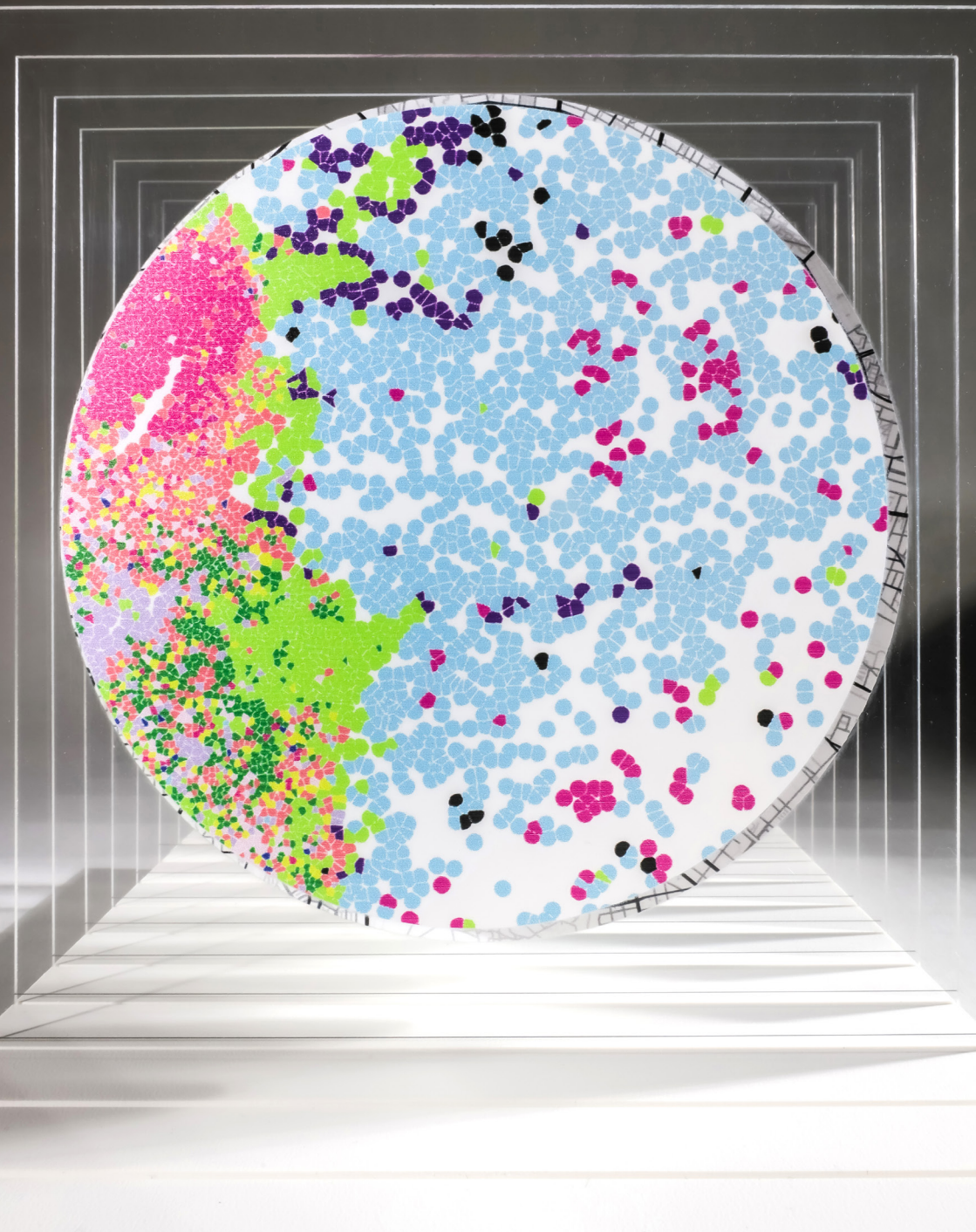
FRIDAY, JUNE 20, 2025

SESSION 6:
Ecosystems in Sequence – Adaptive Genomics from Crops to Coasts
Chair: **Sofia Forslund**

2:00 pm • 2:45 pm	**Jörg Becker,** NOVA University of Lisbon, Portugal *Harnessing male germ line transcriptomics to develop heat-resilient crops*
2:45 pm • 3:00 pm	**Terry Jones,** Charité – Universitätsmedizin Berlin, Germany & University of Cambridge, UK *The recovery of viral genomes from ancient DNA*
3:00 pm • 3:15 pm	**Shuba Varshini Alampalli,** Charité – Universitätsmedizin Berlin, Germany *Living on the Edge: Seasonal Microbial Shifts in Coastal Oceans*
3:15 pm • 4:00 pm	**Detlev Arendt,** EMBL, Heidelberg, Germany *Hotspots of cellular variation*
4:00 pm	**Closing remarks**

SPEAKER ABSTRACTS

SESSION 1



Ulf Landegren

Uppsala University, Sweden

Tools to analyze very many molecules — and even very few

Massive sequencing program have provided comprehensive molecular parts lists of humans and other life forms. My lab is devoted to establishing molecular tools for targeted large-scale analyses of these molecular elements, and for identifying even very rare molecules in the complexity of biological samples for insights in early stages of disease processes. I will describe some of these efforts:

Radically improved opportunities for extensive analyses of dynamic blood biomarkers have created a need for more efficient sample collection. I will describe a solution for self-sampling of blood plasma and cells, and applications for monitoring proteins and antibody specificities.

superRCA is a recent technology that involves a “majority vote” mechanism for highly resolving analyses of miniscule contributions of tumor-derived DNA. We apply the method for monitoring patients via liquid biopsies, providing digital readout of DNA analyses using an installed base of flow cytometers.

For spatial proteomics, multiplex in situ proximity ligation assays can reveal, not just the presence and location of proteins, but also how these proteins are activated to relay cellular information, reflected in dynamic interactions and posttranslational modifications.

In my presentation I will provide a brief historic background for our work and also describe our efforts to make the tools for enhanced biomedical insights broadly available via own and other commercial channels.

Daniel León-Periñán

MDC-BIMSB, Max Delbrück Center, Berlin, Germany

Malva: Real-time Sequence Search across Billions of Cells

Daniel León-Periñán[1], Nikolaus Rajewsky[1], and Nikos Karaiskos[1]

1. MDC-BIMSB, Max Delbrück Center, Berlin, Germany

Modern single-cell and spatial sequencing technologies create increasingly large repositories of nucleotide data from organisms across planet Earth. Current analysis methods require extensive preprocessing, limiting real-time exploration of the sequence-space by researchers and AI agents. Here, we present Malva, a universal search index that unifies multi-terabyte repositories into an instantly queryable, dynamic resource — hundreds of millions of cells across thousands of experiments. Malva enables immediate analysis of any nucleotide pattern with single-cell and spatial resolution, for instance identifying circular RNAs, tracking 3′ UTR dynamics during embryonic development, and detecting alternative splicing events across cell types. Importantly, Malva can discover cell types directly from unaligned reads without reference genomes, allowing cross-species analysis independent of orthology models. The indexed sequences also eliminate preprocessing barriers for AI, directly feeding neural networks that predict spatial expression patterns from primary sequence. By enabling ultrarapid sequence-to-phenotype mapping, Malva opens new avenues for biological discovery in the era of massive sequencing.

Camille Marchet

CNRS, Université de Lille, France

Scaling genomic reuse: hypothesis and algorithms for k-mer collections

The rapid growth of genomic sequencing has created an urgent demand for methods that not only store massive amounts of data efficiently but also support its effective reuse across analyses. At the heart of many such methods lies the concept of k-mer sets, compact representations of sequence content that enable fast querying, indexing, and comparison. This talk will explore recent advances in data structures and algorithms designed to manage large collections of k-mer sets, and will discuss scalability, dynamic updates, and query efficiency. We'll highlight innovations that bridge theory and practice, from algorithmic breakthroughs to real-world applications in areas such as RNA-seq analysis in clinical research.

SPEAKER ABSTRACTS

SESSION 2

Vinicius Maracaja-Coutinho

Advanced Center for Chronic Diseases - ACCDiS, Facultad de Ciencias Químicas y Farmacéuticas, University of Chile, Chile

Addressing inclusion in transcriptomic studies: from animal models to human single-cell biology

Inclusion in transcriptomic research must encompass not only diverse human populations but also the underlying biological variability shaped by sex, cell type, and context. This presentation will explore our integrative approach to advancing equity in transcriptomic studies, spanning both controlled animal models and collaborative efforts within the Human Cell Atlas. At the molecular level, we have employed rodent models to uncover sex-specific transcriptomic responses to stress, with a particular focus on non-coding RNAs. Across three independent studies, our group identified distinct sets of microRNAs (miRNAs) and circular RNAs (circRNAs) that exhibit sex-biased expression and regulatory activity in the prefrontal cortex under stress. These findings revealed not only differential expression patterns, but also sex-dependent regulatory networks and enriched biological pathways, offering mechanistic insights into the higher prevalence of stress-related disorders. Our results underscore the critical need to consider sex as a biological variable in transcriptome-wide studies, especially in the non-coding RNA landscape. Building on this work, we will present for the first time the RatOmics initiative, a new effort to generate a comprehensive, sex- and tissue-specific reference transcriptome and epitranscriptome for Rattus norvegicus. Using long-read sequencing, RatOmics will address longstanding gaps in our understanding of RNA regulation in this essential biomedical model. The project aims to enable more accurate annotations of coding and non-coding RNAs and uncover the dynamics of RNA modifications such as m^6A, m^5C, Ψ, and A-to-I editing across organs and sexes. Complementing this, we will present updates from the LatinCells initiative, a regional consortium dedicated to building single-cell transcriptomic resources across Latin America and the Caribbean. LatinCells focuses on generating and analyzing single-cell data from admixed and indigenous populations in Brazil, Chile, Colombia, Ecuador, Mexico, Peru, Uruguay, and U.S. Caribbeans. By combining inclusive sampling strategies, remote data collection technologies, and locally adapted analysis pipelines, we aim to improve the global representativity of the Human Cell Atlas. Moreover, we highlight our efforts in training, infrastructure building, and community engagement as a model for decentralized genomic science. Together, these three interconnected efforts, sex-aware animal modeling, reference transcriptomics in rat, and population-scale human single-cell biology, underscore how addressing inclusion across species, populations, and molecular layers can lead to a more accurate, equitable, and biologically informed transcriptomic science.

Can Ergen

UC Berkeley, USA
University Hospital Hamburg, Germany

ResolVI – addressing noise and bias in spatial transcriptomics

Can Ergen[1,2], and Nir Yosef[3,1]

1. UC Berkeley
2. University Hospital Hamburg
3. Weizmann Institute of Science

Technologies for estimating RNA expression at high throughput, in intact tissue slices, and with high spatial resolution (spatial transcriptomics; ST) shed new light on how cells communicate and tissues function. A fundamental step common to all ST protocols is quantification, namely segmenting the plane into regions, each approximating a cell, and then collating the molecules inside each region to estimate the cellular expression profile. Despite many advances in this area, a persisting problem is that of wrong assignment of molecules to cells, which limits most current applications to the level of a priori defined cell subsets and complicates the discovery of novel cell states. Here, we develop resolVI, a model that operates downstream of any segmentation algorithm to generate a probabilistic representation, correcting for misassignment of molecules, as well as for batch effects and other nuisance factors. We demonstrate that resolVI improves our ability to distinguish between cell states, to identify subtle expression changes in space, and to perform integrated analysis across datasets. ResolVI is available as open source software within scvi-tools.

Tilman Hoffbauer

Computational Molecular Medicine
Technical University of Munich, Germany

DNA language model based modelling of transcription start sites in twelve yeast species

Tilman Hoffbauer[1], Alexander Karollus[1], Johannes Hingerl[1], and Julien Gagneur[1, 2]

1. Computational Molecular Medicine
2. Institute of Computational Biology

We lack genome annotations for many species, especially for features beyond coding sequences. For example, while the transcription start site (TSS) determines the 5' untranslated region which is crucial for post-transcriptional regulation, its position in non-model organisms is often unclear. Accurate mapping of the TSS across the tree of life is difficult due to the evolution of the regulatory code and lack of experimental data. Recently developed DNA language models from AI trained on increasing numbers of sequenced genomes across the Tree of Life seem promising for tackling this problem, but methods still need to be established. DNA language models can be pre-trained across multiple species from DNA sequence only and then fine-tuned on individual species with in vivo TSS data, leveraging significantly more data than previous models.
Here, we predict CAGE-seq coverage across fungi, spanning 500 million years of evolution. While only using experimental data for 12 species, we fine-tune our language model pre-trained on 800 fungi to achieve remarkable accuracy at predicting the TSS signal for held-out genes (profile pearson correlation coefficient > 0.8 on Saccharomyces cerevisiae), outperforming a BPNet in direct comparison. On held-out species, fine-tuning does improve over a BPNet, but the problem remains hard (pearson < 0.5 on e.g. Candida albicans). The generalization limits of our current model relate to the evolution of a scanning transcription initiation mechanism in yeast and a change in TSS motif preference.
While showing that fine-tuned DNA language models are promising for tasks like TSS prediction, we also highlight that fine-tuning is still bottlenecked by the evolutionary coverage of available experimental data.

Olga Bakina

Max Delbrück Center for Molecular Medicine, Berlin, Germany
Charité – Universitätsmedizin Berlin, Germany

RNA 4sU metabolic labeling in human retinal organoids under dissociation, glucose deprivation and light exposure

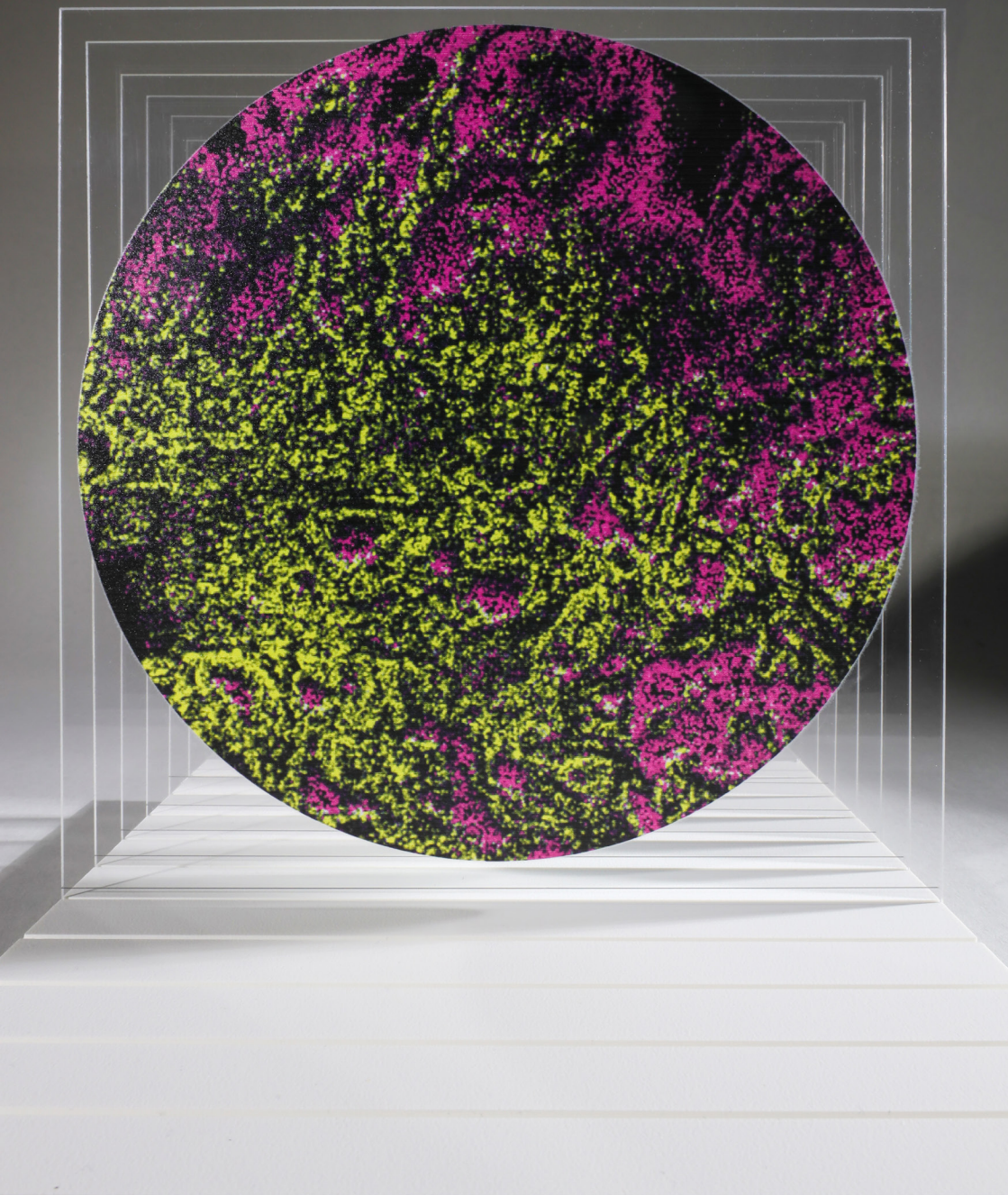
Olga Bakina[1], Anika Neuschulz[1], Gyorgy Bence[2,3], and Jan Philipp Junker[1,4]

1. Max Delbrück Center for Molecular Medicine
2. Institute of Molecular and Clinical Ophthalmology Basel
3. Universität Basel
4. Charité – Universitätsmedizin Berlin

Purpose: RNA sequencing provides us with the possibility to investigate biological states of cells and tissues, albeit with a limitation of accessing a transcriptome in a particular snapshot of time, leaving processes of RNA degradation and synthesis behind the scenes. We have employed a 4-Thiouridine (4sU) metabolic labeling of RNA in human retinal organoids to assess the levels of new RNA synthesis in a defined time period under white light stimulation and single cell tissue dissociation in order to measure the transcriptional response to these stimuli.
Methods: Human retinal organoids were exposed to white light for 4 hrs (Condition 1), additionally the tissue was dissociated to the single cell suspension for 1hr (Condition 2). 4sU was added to the medium to the final concentration of 10-15uM. Afterwards the RNA was extracted from the organoids, sequencing libraries were prepared using a modified Cel-seq2 protocol and samples were sequenced. Sequencing data contains labeled and non-labeled mRNA. Labeled mRNA contains a characteristic T-to-C mutation which is a result of 4sU labeling.
Results: By applying Panther Overrepresentation test to the list of labeled genes, we have identified biological pathways, associated with these genes. Thus, samples dissociated to the single cell suspension included, among others, “Integrated stress response signaling”, “Regulation of DNA-templated transcription in response to stress” and “Intrinsic Apoptotic signaling pathway”. Samples from the white light exposure experiments had following pathways with the highest fold enrichment: “Visual system development”, “Visual perception”, “Sensory perception of light stimulus”
Conclusions: We were able to validate human organoids white light response, as well as stress response activation during tissue dissociation on the RNA level by applying for the first time the 4sU metabolic labelling. Pulse labeling of RNA can be used as a method to access transcriptional changes under short-term exposure of multiple stimuli.

SPEAKER ABSTRACTS

SESSION 3



Helder Nakaya

Hospital Israelita Albert Einstein, São Paulo
University of São Paulo, Brazil

AI for Health: From Precision Diagnostics to Pandemic Preparedness

Artificial intelligence is transforming multiple areas of health sciences, from precision medicine to epidemiology. By integrating molecular data, clinical records, and population-level information, AI enhances diagnostics, disease monitoring, and risk prediction. This talk will explore how AI-driven approaches, including machine learning, computer vision, and generative AI, are shaping personalized treatments, strengthening public health strategies, and advancing our understanding of disease dynamics.

Alexander Lucaci

Weill Cornell Medicine, New York, NY, USA

Diversity and Distinctive Traits of the Global RNA Virome in Urban Environment

Alexander Lucaci[1], Zihao Gao[2], Jun Wu[2], Christopher Mason[1], and Tieliu Shi[2]

1. Weill Cornell Medicine
2. East China Normal University

RNA viruses are the primary catalysts for infectious disease outbreaks, epidemics, and pandemics across multiple hosts including humans. The role of RNA viruses in urban areas remains largely unexplored. This study analyzed the metatranscriptome of 3,326 urban samples from 102 cities in 31 countries, uncovering 84,438 RNA viral units, 85% previously unknown. Two new phyla were discovered, enhancing our understanding of RNA virome phylogenetic diversity. The research also supports the polyphyletic nature of Duplornaviricota and identifies 104 amino acid sites in RNA polymerase that affect virus replication and host interaction. A distinct biogeographical pattern of RNA viruses was observed, indicating potential transmission routes in cities. The study revealed interactions between RNA viruses and antibiotic-resistant ESKAPE pathogens, highlighting urban areas as significant reservoirs for RNA viruses. These findings underscore the need for continuous surveillance and mapping of urban environments to track RNA virus prevalence and dynamics, crucial for public health.

Udo Gieraths

Institute of Virology, Charité – Universitätsmedizin Berlin, Germany
Faculty of Life Sciences

A century-old museum sample reveals a bandavirus with modern day presence in northern European bats

Udo Gieraths[1,2,*], Jörn Beheim-Schwarzbach[1], Matthew J. Pickin[3], Annika Beyer[1,2], Lineke Begeman[4], Bernd Hoffmann[5], Kore Schlottau[5], Martin Beer[5], Rainer G. Ulrich[6,7], Thomas Müller[8], Conrad M. Freuling[8], Tiina Mauno[1], Marco van de Bildt[4], Vera C. Mols[4], Victor M. Corman[1,7], Friedemann Weber[3], Terry C. Jones[1,7], and Christian Drosten[1,7]

1. Institute of Human Genetics, Universitätsklinikum Schleswig-Holstein,
2. Department of Bioengineering and Therapeutic Sciences, University of California,
3. Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA
4. Berlin Institute of Health at Charité - Universitätsmedizin Berlin, 10117 Berlin, Germany
5. Institute of Human Genetics, Universitätsklinikum Schleswig-Holstein,
6. Department of Bioengineering and Therapeutic Sciences, University of California,
7. Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA
8. Berlin Institute of Health at Charité - Universitätsmedizin Berlin, 10117 Berlin, Germany

Ancient genome sequences provide invaluable insights into viral origins and evolution. While ancient DNA viruses have already revolutionized our understanding of viral evolution in humans, ancient RNA viruses may similarly reshape our perspective on zoonotic viruses within animal reservoirs. Here we explored the feasibility of detecting ancient viral RNA in ethanol-preserved bat samples from a museum collection. We successfully recovered the coding-complete genome of a bandavirus, a negative-sense segmented RNA virus that clusters with the highly pathogenic human severe fever with thrombocytopenia syndrome virus (SFTSV). The virus was detected in a Common pipistrelle (Pipistrellus pipistrellus) bat collected in northern Germany in 1919, making it one of the oldest sequenced mammalian pathogenic RNA viruses, only comparable to historic Measles and Influenza A viruses from 1912 and 1918.

>>>

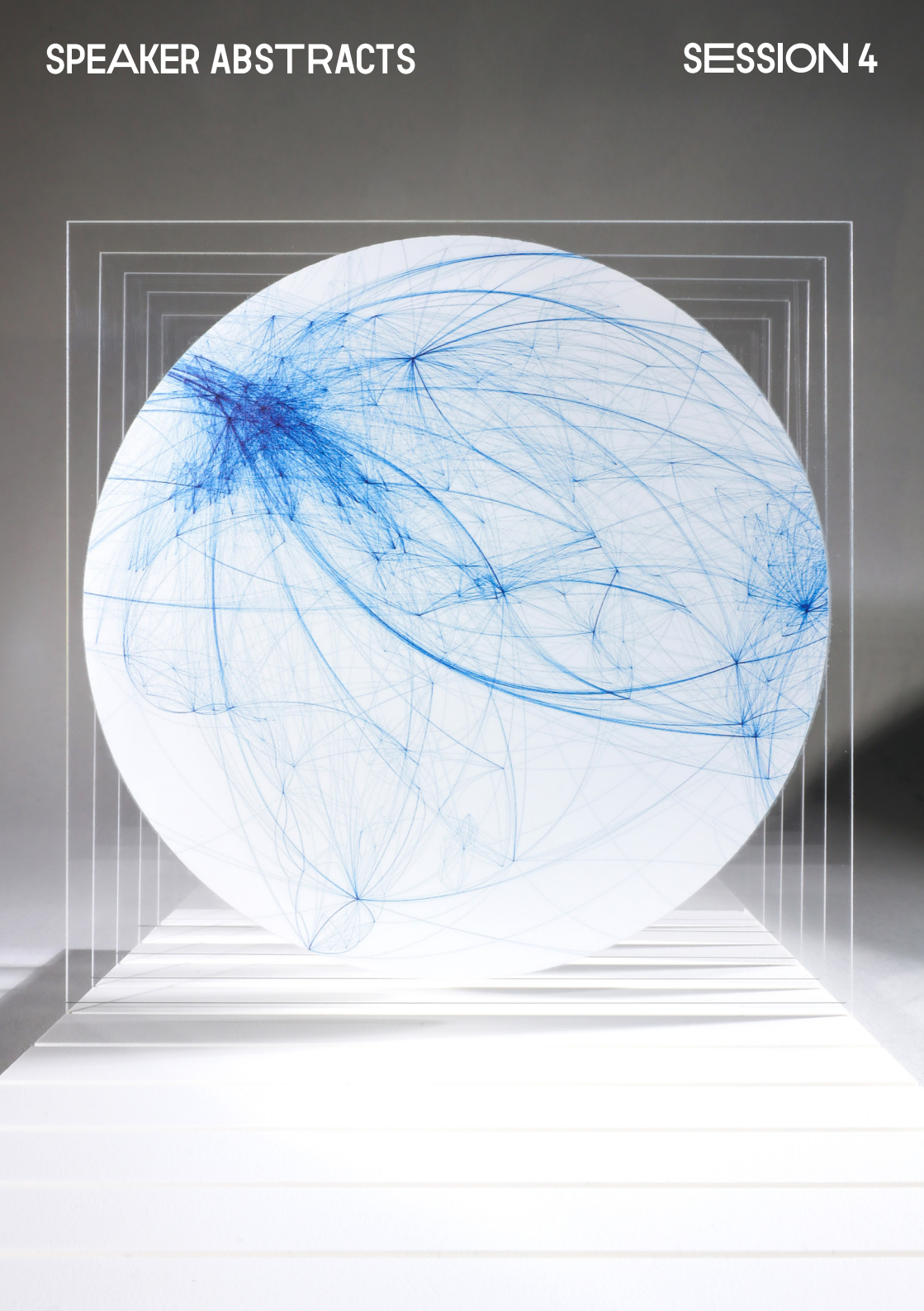
>>> Screening 1086 contemporary bat samples revealed strains of the same virus species in nine Common Pipistrelle bats (2010–2018) and one Serotine bat (Eptesicus serotinus, 1999) from Germany and the Netherlands. Coding-complete genomes indicate frequent genome segment reassortment and widespread circulation of reassortants of this understudied virus species (Bandavirus zwieselense). Although initial analysis suggested a temporal signal, further tests failed to confirm it, raising concerns about the feasibility of other dated phylogenetic analyses for Bandaviruses that are based on much narrower sampling windows. Additionally, functional assays demonstrated that the virus's nonstructural (NSs) protein effectively inhibits interferon induction in human HEK-293T cells.
Our findings highlight the feasibility and scientific value of extracting and analyzing ancient viral RNA from ethanol-preserved museum specimens to substantially enhance our understanding of RNA virus evolution.

Marion Koopmans

Erasmus Medical Center, Rotterdam, the Netherlands

SPEAKER ABSTRACTS

SESSION 4



Beate Kampmann

Charité Center for Global Health
Charité - Universitätsmedizin Berlin, Germany

A systems-biology approach to unravel the ontogeny of the immune system in newborns and the impact of vaccination

This presentation introduces a large scale study in newborn babies, conducted in a resource-poor setting that investigated the impact of vaccines given at birth on pathways of neonatal immune development in unprecedented detail. I will summarise the findings from our multi-omic approach and put the data into context of types of vaccines, maternal antibody, immune ontogeny, sex and geography and share with you, how much we continue to learn from a Small sample -Big data approach.

Natalia von Staa Mansur

School of Pharmaceutical Sciences - University of São Paulo (USP), Brazil

Graph Neural Networks in Precision Medicine: Predicting Clinical Outcomes of Infectious Diseases

Natalia von Staa Mansur[1], and Helder Nakaya[2,3]

1. School of Pharmaceutical Sciences - University of São Paulo (USP)
2. University of Sao Paolo
3. Hospital Israelita Albert Einstein

Understanding how gene expression shapes the immune response to infectious diseases is essential for uncovering mechanisms of pathogenesis, identifying therapeutic targets, and improving diagnostic strategies. Transcriptomic signatures identified through machine learning (ML) techniques are increasingly being explored to support precision medicine. However, the high dimensionality and sparsity of transcriptomic data often lead to overfitting and reduced model stability. It is suggested that considering gene relationships may improve prediction accuracy in biological data. This project aims to explore the use of graph neural networks (GNNs) to model gene expression data within the context of gene-gene relationships, aiming to develop more robust and accurate classifiers for infectious disease outcomes. We focus on Mycobacterium tuberculosis infection, a major global health challenge. Current tuberculosis (TB) diagnostics lack sensitivity and often fail to identify individuals with latent infection who are at high risk of progressing to active disease. RNA-seq data from individuals with varying TB statuses were analyzed, and differentially expressed genes were identified between progressors and non-progressors, as well as between latent and active TB cases. A GNN-based binary classifier is currently being trained to predict clinical outcomes based on both expression profiles and gene network topology. Preliminary results using progressor and non-progressor expression data and an initial PPI matrix reached 95.45% test accuracy, 93.70% validation accuracy, 95.43% test F1-score, and 93.36% validation F1-score. Next steps involve refining the biological network, by integrating selected genes using publicly available protein–protein interaction and co-expression data. Further on, it will be improved model generalization, and applied layer-wise relevance propagation to interpret predictions. The most predictive genes will be functionally analyzed and mapped onto interaction networks to identify key regulatory elements. This approach supports the development of precision diagnostics and deepens our understanding of host-pathogen interactions in TB and other infectious diseases.

Grace Androga & Brenda Kwambana-Adams

Malawi Liverpool Wellcome Programme, Blantyre, Malawi
College of Medicine, University of Ibadan, Ibadan, Nigeria
Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

Genomics without borders to inform vaccine deployment against outbreak-prone bacteria in Africa

Grace Opia Androga[1,2], Brenda Kwambana-Adams[1,3]

1. Malawi Liverpool Wellcome Programme, Blantyre, Malawi
2. College of Medicine, University of Ibadan, Ibadan, Nigeria
3. Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

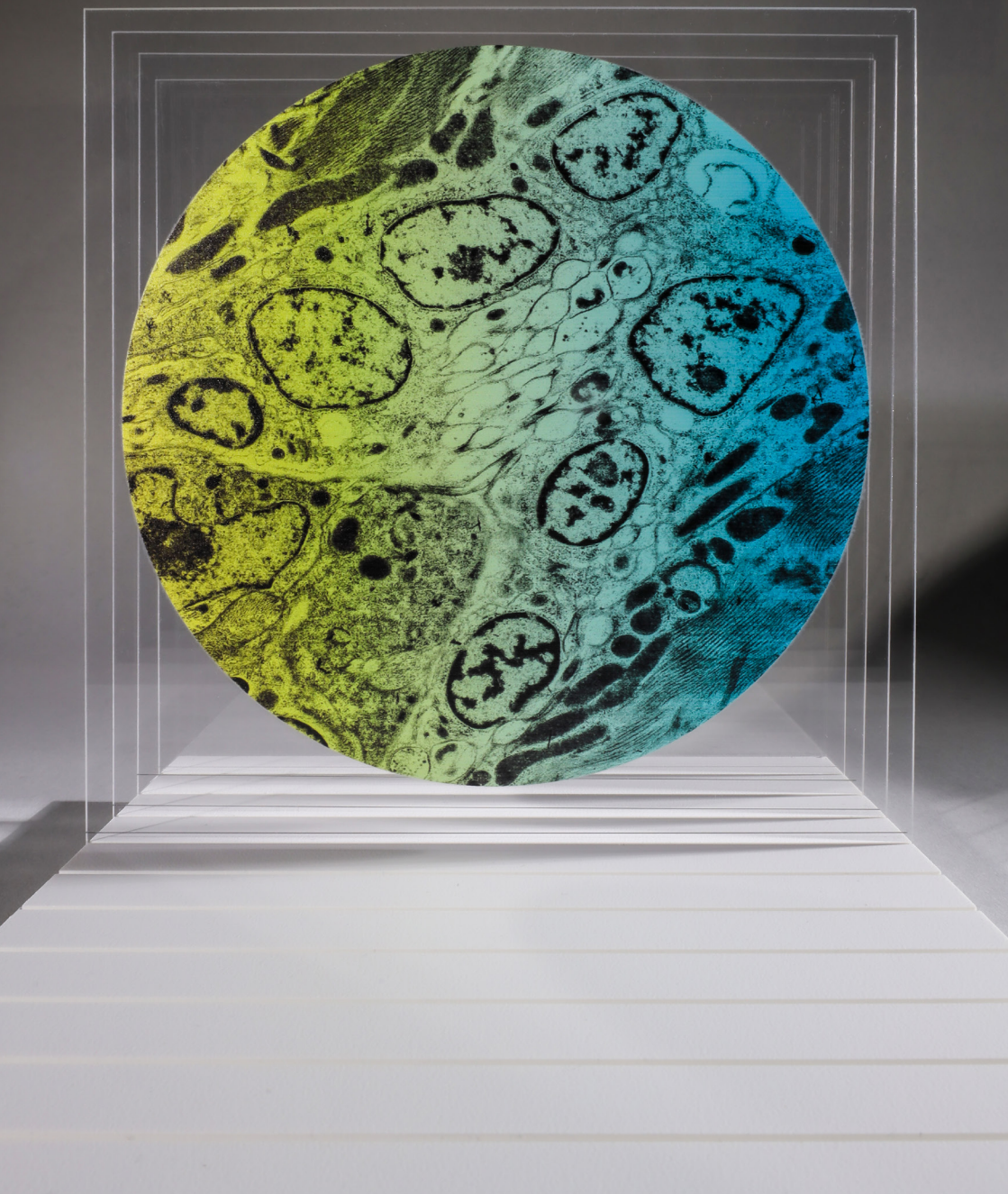
Genomics-informed surveillance is revolutionising infectious disease control in high-income settings, enabling earlier outbreak detection, targeted interventions, and data-driven vaccine strategies. However, as the antimicrobial resistance (AMR) crisis intensifies and the threat of infectious disease outbreaks persists, there is a pressing need to extend these capabilities to low- and middle-income countries through a “genomics without borders” approach. This demands equitable access to sequencing technologies, molecular diagnostics, and the data ecosystems that inform public health action.

This talk will present our efforts to integrate high-resolution genomics with epidemiological surveillance to characterise the invasive disease and transmission dynamics of vaccine-preventable and soon-to-be vaccine-preventable bacterial pathogens with outbreak potential. We leverage portable nanopore sequencing and novel CRISPR-based diagnostics to develop field-adaptable, culture-independent tools for detecting multi-drug-resistant bacteria and fastidious organisms, even at low densities.

These innovations are closing critical diagnostic gaps and enabling earlier detection of vaccine escape variants and emerging bacterial threats, thereby strengthening regional and global preparedness against AMR and epidemic-prone diseases in settings with the greatest gaps.

SPEAKER ABSTRACTS

SESSION 5



Alfred Ngwa

MRC Unit, The Gambia, London School of Hygiene & Tropical Medicine, UK

Evolutionary dynamics and malaria transmission in Africa

Jonathan Wood

Wellcome Sanger Institute, Hinxton, UK

Tree of Life – generating reference genomes at scale.

High-quality, chromosomally complete genome assemblies are an essential starting point for developing new model systems. Recent advances in sequencing technology have dramatically accelerated the availability of reference genomes across the tree of life, spearheaded by large-scale sequencing projects.
The Darwin Tree of Life project is part of a wider global project to sequence the genomes of all eukaryotic life, known as the Earth BioGenome Project (EBP). The DToL project aims to generate chromosomally resolved genome assemblies to EBP standards for all described species in the British and Irish Isles and to date have released over 2,500 genome assemblies. These resources underpin discovery science and enable lab and field-based researchers to gain new insights into their study systems.

Here we present the pipelines and processes that enable the generation and release of DToL reference genomes at scale.

Tobias Marschall

Institute for Medical Biometry and Bioinformatics, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University, Düsseldorf, Germany
Center for Digital Medicine, Heinrich Heine University Düsseldorf, Germany

Pangenome based analysis of structural variation

Breakthroughs in long-read sequencing technology and assembly methodology enable the routine de novo assembly of human genomes to near completion. Such assemblies open a door to exploring structural variation (SV) in previously inaccessible regions of the genome. The Human Pangenome Reference Consortium (HPRC) and the Human Genome Structural Variation Consoritum (HGSVC) have produced high quality genome assemblies, which provide a basis for comparative genome analysis using pangenome graphs.

First, we will ask how a pangenomic resource like this can be leveraged in order to better analyze structural variants in samples from large cohorts with short-read whole-genome sequencing (WGS) data. In a process called genome inference implemented in the PanGenie software, we can use a pangenome reference to infer the haplotype sequences of individual genomes to a quality clearly superior to standard variant calling workflows. This process allows us to detect more than twice the number of structural variants per genome from short-read WGS and therefore provides an opportunity for genome-wide association studies to include these SVs.

Second, we introduce Locityper, a tool specifically designed for targeted genotyping of complex loci using short and long-read whole genome sequencing. For each target, Locityper recruits and aligns reads to locus haplotypes and finds the likeliest haplotype pair by optimizing read alignment, insert size and read depth profiles. Locityper accurately genotypes up to 194 of 256 challenging medically relevant loci (95% haplotypes at QV33), an 8.8-fold gain compared to 22 genes achieved with standard variant calling pipelines. Furthermore, Locityper provides access to hyperpolymorphic HLA genes and other gene families, including KIR, MUC and FCGR.

SPEAKER ABSTRACTS

SESSION 6

Jörg D. Becker

Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa (ITQB NOVA),
Av. da República, 2780-157 Oeiras, Portugal

Harnessing male germ line transcriptomics to develop heat-resilient crops

Climate change, with its rising temperatures and increasingly common heatwaves, poses a major risk to the yields of fruits and seeds vital to human diets. Heat stress has a particular impact on the development of male/female gametes and fertilization in flowering plants. In our group, we focus on the impact during the progamic phase, when a pollen tube grows through the pistil to deliver two sperm cells for double fertilization. We combine Fluorescence-activated cell sorting of sperm cells from the model plant Arabidopsis thaliana and the crop Solanum lycopersicum (tomato) with ultra-low input transcriptomics and kingdom-wide gene expression analyses. Through this approach and further characterization of the identified candidate genes, we aim to enhance the resilience of plant fertilization processes under climate stress conditions.

Terry Jones

Charité Universitätsmedizin
University of Cambridge

The recovery of viral genomes from ancient DNA

The first viral genomes were obtained from contemporary samples / infections (MS2 RNA phage in 1976, DNA phage φX174 in 1977, then EBV in 1984, VZV in 1986, etc.). Later, complete viruses genomes were identified in the host genomes of various eukaryotes and bacteria, and even in other viruses. Most recently, viral genomes have been inferred from ancient DNA, and a few ancient viruses have been resurrected. These three sources of viral genomes offer quite different information and sometimes conflicting perspectives on viral evolution. Viral genome recovery from ancient samples is a field still its infancy, regularly producing tantalizing if sporadic results, but with an unclear future. This talk will cover four areas. First, a summary of findings of RNA and DNA viral genomes from sample material between 100 and 50,000 years old. Then, description of the novelties (e.g., extinction, gene loss, genome insertions and deletions, geographic movement) and potential upsides of these investigations (e.g., how the past may inform pharmaceutical and surveillance measures). Third, some cautions regarding interpretation (e.g., algorithmic assumptions, substitution rates, impact of under-sampling, ancient prevalence and pathogenicity). Finally, some thoughts regarding future prospects for the recovery of viral genomes from ancient materials.

Shuba Varshini Alampalli

Charité - Universitätsmedizin Berlin, Germany

Living on the Edge: Seasonal Microbial Shifts in Coastal Oceans

Shuba Varshini Alampalli[1], Stefanie Böhnke[2], Nicole Adam-Beyer[2], Klaus Wallmann[2], Katja Laufer-Meiser[2], Emanuel Wyler[3], and Mirjam Perner[2]

1. Charité - Universitätsmedizin Berlin
2. GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel
3. Max-Delbrück-Centrum für Molekulare Medizin (MDC)

Coastal deoxygenation is intensifying under climate change, with the Baltic Sea responding rapidly to stratification, eutrophication, and nutrient loading. At the seasonally hypoxic Boknis Eck site (southwestern Baltic Sea coastal waters, at 28m water depth), we analyzed microbial functional dynamics in surface sediments (0–2 cm) across four seasonal time points in 2022 (February, March, early October, late October), integrating metagenomes, metatranscriptomes, and sediment–water column chemistry.

Eighty-three high-quality metagenome-assembled genomes captured diverse sulfur-cycling Desulfobacterota, organic-matter degrading Planctomycetota, and sulfur-oxidizing Beggiatoaceae. Microbial transcriptomes revealed distinct seasonal shifts aligned with redox gradients and organic matter inputs linked by phytoplankton blooms. Early October, following a phase of anoxic bottom waters and during deep-water reoxygenation, exhibited peak expression of NAD, thiamine, nucleotide biosynthesis, gluconeogenesis, and glycogen degradation—indicating anabolic reprogramming, redox balancing, and carbon recycling from C1–C5 intermediates. This coincided with rising bottom water oxygen (about 30 µM), increased CH_4 emissions from the sediments, sulfide accumulation in sediments, and peak organic matter degradation rates.

>>> In contrast, late October showed enrichment of acetate oxidation, anammox, fatty acid, and deoxyribonucleotide biosynthesis, corresponding with residual anoxia, high sulfate reduction, sulfide buildup, and modeled pyrite formation. Despite bottom water O_2 reaching about 178 µM, surface sediment metabolism remained dominantly anaerobic.

Winter (Feb–Mar) transcriptomes reflected aerobic heterotrophy, with upregulated glycolysis, Calvin cycle, formaldehyde assimilation, and fatty acid synthesis, aligned with fully oxygenated bottom waters, high nitrate, and organic matter input. Vancomycin resistance and other stress pathways were enriched, possibly reflecting competitive bacterial dynamics.

These findings highlight early October as a transitional metabolic hotspot, with maximal microbial flexibility and niche restructuring. Seasonal oxygenation pulses dynamically couple redox-sensitive carbon, nitrogen, and sulfur cycling in coastal sediments. Our results demonstrate how sequencing-based approaches can enhance understanding of biogeochemical change in hypoxic marine environments.

Detlev Arendt

Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany
Centre for Organismal Studies, University of Heidelberg, Germany

Hotspots of cellular variation

Planetary biology addresses the interplay of genomes and phenomes in ecosystems at all levels of resolution, from single-celled unicellular to multicellular organisms, holobionts, and organismal communities, towards an AI-driven imaging-sequencing loop across scales.

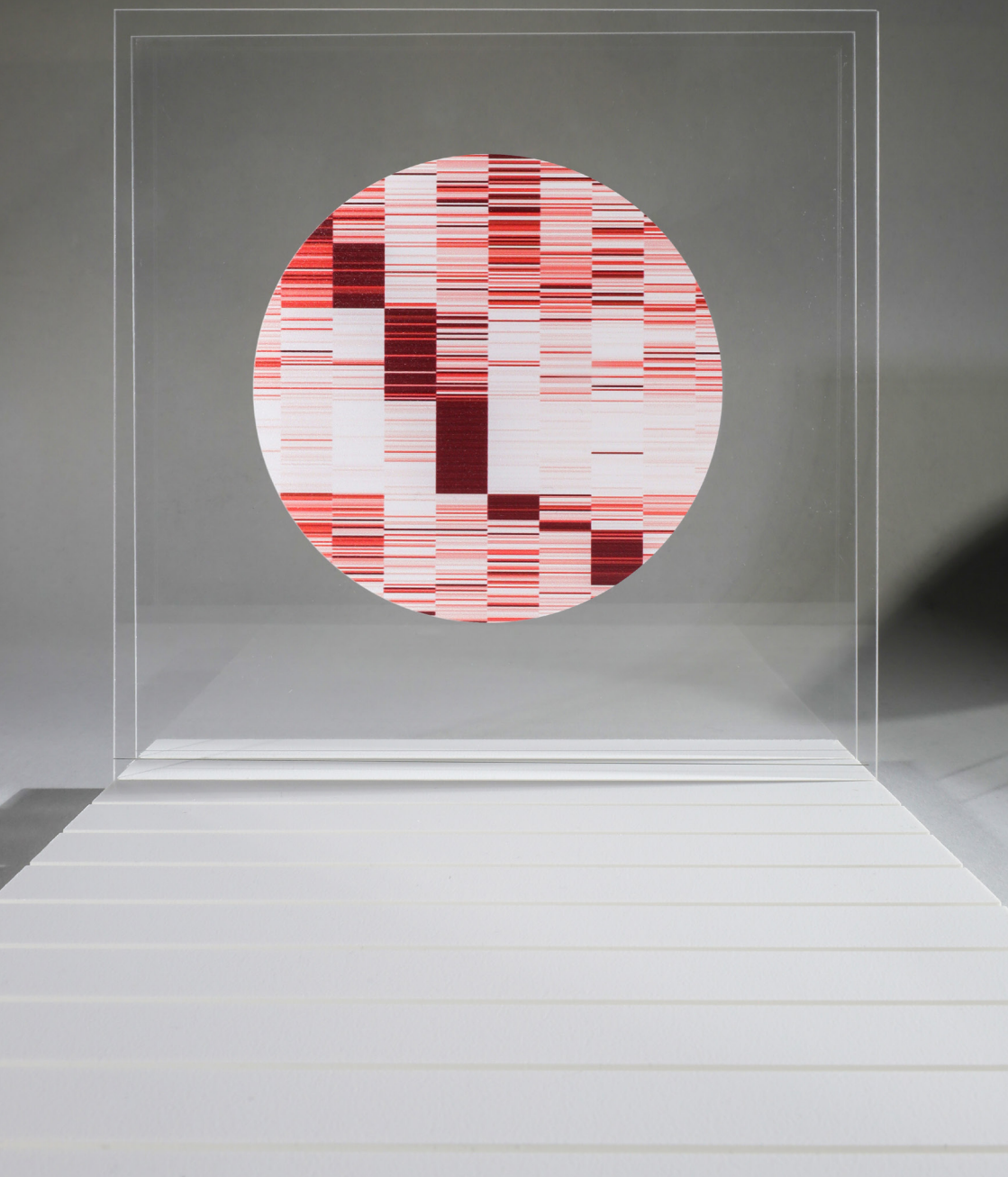
Towards this aim, we are targeting the genotype-phenotype link in the marine annelid Platynereis dumerilii. Taking advantage of its highly stereotypic development, we have built the PlatyBrowser, which combines cellular expression atlases and volume electron microscopy to establish a link between gene expression and cellular and subcellular morphologies for all cell types of the six-days-old young worm.

Mapping a whole-body single-nuclei expression atlas onto the PlatyBrowser, we have identified more than 270 cell types. Remarkably, the non-neuronal cell types are specified by a handful of highly conserved gene expression programs with core regulatory complexes controlling the expression of large sets of coregulated genes, called coregulons. These represent families of related cell types that together form tissues and organs and represent the main unit of evolutionary change.

Building on this multimodal atlas, we are now aiming at identifying cellular hotspots of variation that drive environmental adaptation. For this, we correlate morphological variation that we obtain from cellular-resolution X-ray of 6dpf young worms with different modes of variation in single-cell expression data for all cell types in animals collected from multiple sampling sites during the “Traversing Ecosystems” (TREC) expedition along European coastlines.

POSTER ABSTRACTS

In alphabetical order



Transcriptome Analysis and Validation from FFPE Tissue Identifies Stage-Specific Gene Expression Profiles Differentiating Adenoma, Carcinoma In-Situ and Adenocarcinoma in Colorectal Cancer Progression

Reem Alhamidi[1,*], Faisal Alhosani[1], Burcu Ilce[1], Alaa Altaie[1], Nival Ali[1], Alaa Hamad[1], Axel Künstner[2], Cyrus Khandanpour[3], Hauke Busch[4], Basel Al-Ramadi[5], Rania Harati[1], Kadria Sayed[6], Ali AlFazari[7], Riyad Bendardaf[8], and Rifat Hamoudi[9]

1. University of Sharjah - Medical Campus, Research Institute of Medical and Health Sciences
2. Medical Systems Biology Group, Lübeck Institute of Experimental Dermatology, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
3. University Cancer Center Schleswig-Holstein, University Hospital Schleswig-Holstein, University of Lübeck, 23562 Lübeck, Germany
4. Lübeck Institute of Experimental Dermatology, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
5. College of Medicine and Health Sciences, United Arab Emirates University, Al Ain P.O. Box 15551, United Arab Emirates
6. American Hospital Dubai, Dubai P.O. Box 3050, United Arab Emirates
7. Mediclinic Welcare Hospital
8. Oncology Unit, University Hospital Sharjah, Sharjah P.O. Box 72772, United Arab Emirates
9. Division of Surgery and Interventional Science, University College London, London WC1E 6BT, UK*

* Corresponding author

Colorectal cancer (CRC) progresses through three stages: adenoma, carcinoma in situ (CIS), and adenocarcinoma. Despite therapeutic advances, the absence of reliable stage-specific biomarkers limits accurate diagnosis and targeted treatment. This study aimed to identify transcriptomic signatures and molecular pathways specific to each CRC stage. We conducted RNA sequencing on FFPE tissue samples from adenoma (n = 10), CIS (n = 8), and adenocarcinoma (n = 11), followed by absolute gene set enrichment analysis (absGSEA). Key findings were validated using RT-qPCR and additionally validated in an independent clinical cohort of 1926 CRC patients, including survival analysis for 1336 cases.

Pathway analysis revealed distinct stage-specific mechanisms. CIS samples showed enriched apoptotic and Wnt signaling pathways compared to adenoma. Adenocarcinoma samples were enriched for

>>> transcriptional co-regulation, protein kinase activity, and extracellular matrix (ECM) organization compared to earlier stages. Several genes showed stage-specific expression: ARRB1, CTBP1, and CTBP2 in adenoma; RPS3A and COL4A5 in CIS; and COL1A2, CEBPZ, MED10, and PAWR in adenocarcinoma. Functional analysis confirmed ARRB1 and CTBP1/2 roles in early tumorigenesis, while COL1A2 and CEBPZ were linked to ECM remodeling and transcriptional regulation in advanced CRC.

Experimental RT-qPCR validation confirmed the differential expression of ARRB1, RPS3A, COL4A5, COL1A2, and MED10 across stages, supporting their utility as stage-specific biomarkers. These findings provide a comprehensive molecular framework for CRC progression, offering potential diagnostic and prognostic biomarkers that could inform stage-tailored therapeutic strategies.

Spatial Organization and function of RNA in primary neuroblastomas and their TME

Florian Bartsch[1, 2,*], Svea Beier[2, 3], Angelika Eggert[1], and Nikolaus Rajewsky[2, 4]

1. Charité
2. Max-Delbrück Center for molecular medicine / Berlin institute of medical systems biology
3. Humboldt Universität zu Berlin
4. Charité - Universitätsmedizin Berlin

* Presenting author

The tumor microenvironment (TME) of neuroblastoma is a dynamic ecosystem of diverse cells that interact with each other leading to high inter- and intra-tumoral heterogeneity. Spatial architecture within the neuroblastoma TME shapes tumor evolution and therapeutic response. Using a novel high resolution spatial transcriptomic technique (Open-ST), we aim to study spatial patterns of these TMEs of patient tissue samples in low- and high-risk tumors. We aim to computationally predict effects of spatial TME remodeling and tumor plasticity across risk status. To achieve this, we generated a high-throughput, high-resolution spatial transcriptomics dataset of more than 10 patient tumor samples, capturing about 1 million cells. In order to identify shared- and non-shared features, we initially decomposed gene expression data into gene modules using Non-Negative-Matrix factorization (NMF). Spatial coordinates from combined high- and low-risk samples are leveraged to train a neural-network (Multi-layer perceptron (MLP)) to learn spatial organization of gene module activites. Preliminary results show that the model captures module activity in space and can be used to predict gene programs across risk status in space. By mapping modules on cell-types and studying their spatial organization, novel mechanisms driving TME remodeling are resolved. This data-driven strategy aims to unravel mechanisms driving spatial architecture and cellular composition of neuroblastoma, shedding light on tumor heterogeneity and disease progression. Ultimately, this can lead to the identification of novel clinical markers.

Mapping neuroblastoma tumor and microenvironment heterogeneity with cellular resolved spatial transcriptomics

Svea Beier[1, 2,*], Tancredi Massimo Pentimalli[1, 3], Katrin Banko[3], Ilan Theurillat[1], Angelika Eggert[3], and Nikolaus Rajewsky[1, 3]

1. Max-Delbrück Center for Molecular Medicine / Berlin Institute of Medical Systems Biology
2. Humboldt Universität zu Berlin
3. Charité – Universitätsmedizin Berlin

* Presenting author

Neuroblastoma is a rare childhood malignancy arising in the developing sympathetic nervous system, and responsible for approximately 10% of all childhood cancer-related deaths. This disease is remarkably heterogeneous — ranging from spontaneous regression in infants, and excellent survival in low-risk patients to often lethal courses in high-risk cases. Tumor heterogeneity is strongly linked to therapeutic outcome, making it essential to understand the molecular mechanisms that drive this variability.

While the tumor microenvironment plays a key role in tumor progression, its contribution to neuroblastoma biology remains poorly understood. To address this gap, we apply Open-ST, a high-resolution and cost-effective spatial transcriptomic method, to analyze neuroblastoma samples across risk groups. So far, we sequenced almost one million spatial spots from 9 patients, representing the beginning of our growing cohort.

Here we apply GeneNMF, a non-negative matrix factorization-based framework, to identify shared and sample-specific metaprograms from spatial data. The identified shared and sample-specific metaprograms represent different cell type identities as well as cancer cell states, that can be linked back to published single-cell RNA-seq references. By this, we achieve harmonized cell type annotation across multiple samples, allowing even the detection of rare populations such as neutrophils and mast cells at high resolution. This analysis reveals patient differences and treatment effects in immune and malignant organization, underscoring the diversity of the neuroblastoma microenvironment and cellular plasticity.

This work lays the foundation for future analysis aimed at decoding the microenvironmental mechanism behind neuroblastoma heterogeneity, with the goal of advancing diagnosis and therapeutic approaches.

Deletion of ATRX in Mature Hippocampal Neurons is Responsible for Chromatin Dysregulation linked to Neuronal Dysfunction

Berta J-A Hacha[1, 2,*], Izabela-Cezara Harabula[1], Alexander Kukalev[1], Dominika Vojtasova[1, 3], Julia Tiffany Brott[4, 5], Renee J Tamming[4, 5], Lucia Peixoto[6], Nathalie Bérubé[4, 5], and Ana Pombo[1, 3]

1. Max Delbrück Center for Molecular Medicine, Berlin Institute for Medical Systems Biology, Hannoversche Str. 28, 10115 Berlin, Germany.
2. Technische Universität Berlin, 10623, Berlin, Germany
3. Humboldt-Universität zu Berlin, 10117 Berlin, Germany.
4. Western University, London, ON, Canada
5. Children's Health Research Institute, London, ON, Canada
6. Washington State University, Sleep and Performance Research Center, Elson S. Floyd College of Medicine

* Presenting author

Chromatin regulators have broad functions in most cell types but are often mutated in association with complex neurodevelopmental disorders that are accompanied by a spectrum of neurological impairments. In-vivo conditional knockout of the chromatin remodeler ATRX (Alpha thalassemia/mental intellectual disability syndrome X-linked) in post-mitotic glutamatergic neurons results in strongly impaired hippocampal-related learning and long-term memory in fear-based experiments in mouse models. To examine the genomic targets and mechanisms of altered gene expression induced by ATRX depletion in adult pyramidal glutamatergic neurons (PGNs) of the hippocampal CA1, important for memory, we have applied Genome Architecture Mapping (GAM) and single-cell analysis of gene expression and chromatin accessibility, using 10x Genomics Multiome snRNA-seq/snATAC-seq. Conditional deletion of ATRX in PGNs results in changes in compartment A (euchromatin) and B (heterochromatin) with a tendency to euchromatinization and a different set of Topologically Associated Domains (TADs) that identify lost and de novo TAD borders upon ATRX deletion. Current analyses using Multiome 10X snRNA-seq and snATAC-seq aim to determine changes in gene expression and chromatin accessibility specifically in mature PGNs of the CA1 region, in WT and mutant genotypes, to further unravel the chromatin-based mechanisms that are deregulated by ATRX that cause extensive neural dysfunction.

DEGDB: A Comprehensive Workflow for Integrating Geospatial Data in Biomedical Research

Marcel Jentsch[1,*]

1. BIH
* Presenting author

Environmental factors are increasingly recognized as influential determinants of human health, contributing to a wide range of physical and psychological challenges. Integrating environmental data with clinical and genomic information enables a more comprehensive understanding of health and disease, allowing researchers to identify relations and patterns that might otherwise remain undetected. The Deep-Envirotyping-Geo-Database (DEGDB) framework is designed to generate a comprehensive and accessible geospatial database, along with tools to work with this database. DEGDB overcomes fragmentation of geospatial data by supporting the downloading, harmonization, and integration of geospatial data from multiple sources into a unified database. This framework enables the integration of geospatial data into health research by using geolocation to link environmental profiles with health data ascertained in existing observational studies. DEGDB supports continuous updates and expansions, ensuring that the database remains current and relevant to ongoing research efforts, while saving time and resources and enhancing data accuracy and reliability. By harmonizing publicly available data from various sources into a unified database that adheres to the FAIR (Findable, Accessible, Interoperable, and Reusable) criteria, DEGDB enhances accessibility and promotes reproducibility in biomedical research.

Identification and Mapping the 3D Spatial Organisation of Bacterial Species During Plant Fiber Digestion Using MetaGAM

Simon Middeldorf[1, 2,*], Luiz Gustavo dos Anjos Borges[1], Alexander Kukalev[1], Sergej Herzog[1], Lotte Victoria Winther Stagsted[1], Itai Amit[3], Sarah Moraïs[3], Itzhak Mizrahi[3], and Ana Pombo[1, 4]

1. Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Berlin, Germany
2. Humboldt-Universität zu Berlin, Berlin, Germany
3. Ben-Gurion-University of the Negev, Beer-Sheva, Israel
4. Johns Hopkins University, Baltimore, MD, USA

* Presenting author

The digestion of plant fiber cellulose in the gut of animals, including humans, is an enzymatic process commonly associated with bacteria. Understanding the microbial interactions and their impact on the host is key in advancing optimal personalized nutrition, association with disease, and livestock production. High-throughput sequencing methods are revolutionizing the identification of microorganisms across samples with varying complexity. However, achieving sufficient sensitivity and spatial resolution at the same time remains challenging. Using our newly developed method metaGAM, we seek to investigate the spatial association between fiber-degrading bacterial communities and nutritional plant particles, uncovering the intricate relationships during microbial cellulose degradation.

Here, we apply metaGAM, to investigate the dominance of specific bacterial species in a model of plant fiber degradation. MetaGAM is a high-throughput sequencing method that sensitively detects microbial species and measures preferred microbial spatial arrangements, by physical slicing of three-dimensional microbial communities in their ecological environment combined with whole genome sequencing and spatial statistics. We focused on a bacterial consortium of nineteen bacteria species which was cultivated for 72 hours in vitro on grinned plant fibers of Triticum aestivum, one of the most widely used wheat varieties in livestock nutrition globally. After gelatin embedding of the resulting microbiota-plant fiber consortia, we created ultrathin cryosections (~ 0.3 µm thick) and precisely microdissected plant particle landscapes. By extracting and amplifying the total genetic content

>>> of individual microsamples, each with an approximate total volume of 3000 µm3, it was possible to characterize bacterial consortia surrounding nutritional particles.

We successfully identified the bacterial strains across the collection of samples. Our spatial analyses detected preferred spatial associations among a cluster of eleven species, while certain species appear to exclude each other. Taken together, our data suggest that bacterial species compete for dominance within the microbiome associated with the digestion of plant fibers.

Unravelling the clonal dynamics of somatic mutations to learn mechanisms of selection in human disease

Benedict Monteiro[1, 2,*], Martina Macino[1, 2], Soniya Shende[1, 2], Patrick Weidner[1, 2], Oliver Dyck Dionisi[1, 2], and Ashley Sanders[1, 2]

1. Max Delbrück Center for Molecular Medicine
2. Berlin Institute of Health

* Presenting author

To date a wide range of somatic mutations have been detected in the genomes of human disease samples, including single nucleotide variants (SNVs), small insertions and deletions (indels) and large-scale structural variations (SVs). The clonal prevalence of somatic mutations is dynamic throughout disease progression, which can be used to make biological inferences of the selective pressures acting on the affected tissue. However, to our knowledge there is no method to characterise the complete spectrum of somatic mutations and its clonal evolution in the same sample.

To address this, we are developing computational frameworks to identify all classes of somatic mutation in data from Strand-seq, a haplotype-resolved single strand DNA sequencing technique that has greater power to resolve SVs than conventional methods. As somatic SNVs and indels have not previously been studied in Strand-seq data, we are currently assessing our powers of detection by benchmarking our calls against orthogonal datasets. Preliminary results suggest that we are able to capture a sufficient number of somatic mutations that can be used for further study.

We are then leveraging mutational profiles to reconstruct the phylogenetic relatedness of both subclones and individual cells, which will allow us to infer the contribution of different mutation types to the clonal architecture of inflammatory bowel disease and cancer samples. By integrating new variant classes into our analyses, we expect to gain new insights into how the selective pressures experienced by a tissue influence disease development and progression.

Inferring a Cross-species Atlas of Regulatory Elements from Genomic Language Models spanning 500 Million Years of Evolution

Johann Promeuschel[1,*], and Julien Gagneur[2, 3]

1. Technical University of Munich
2. Computational Molecular Medicine
3. Institute of Computational Biology

* Presenting author

Non-coding genomic regions contain functional regulatory elements that are major determinants for gene expression. Those include transcription factor binding sites as well as post-transcriptional regulatory elements in the RNA. Still, a comprehensive understanding of the regulatory grammar connecting these elements is lacking, an issue that is exacerbated in species outside of non-model organisms where experimental data is scarce.

Genomic language models (gLMs) trained on a vast body of genomic sequences covering large parts of the tree of life have been shown to allow for the discovery of functional elements and their dependencies, taking into account sequence and species context. However, their potential to map functional elements and describe their evolutionary trajectories across long periods of time remains mostly unexplored. Here, we took advantage of the recently described gLM-informed nucleotide dependency maps to identify and cluster potential functional elements 1 kilobase upstream of start codons of protein coding genes in 13 fungal species covering more than 500 million years of evolution. We discovered over 100,000 sites grouped into 559 motifs. Among them we identified well known recognition motifs for fungal DNA binding factors such as RAP1 as well as multiple clusters of motifs with low frequency in individual species, but that are reappearing across species. The motif atlas will be continuously expanded to new species with the aim to guide functional validation and accelerate annotation of regulatory elements especially in non-model organisms.

Multiple instance fine-mapping: predicting causal regulatory variants with a deep sequence model

Alexander Rakowski[1,*], and Christoph Lippert[1, 2]

1. Hasso Plattner Institute
2. Hasso Plattner Institute for Digital Health at Mount Sinai

* Presenting author

Identifying causal genetic variants in a computational manner remains an open problem. Training end-to-end prediction models is not possible without large ground-truth datasets, while genome-wide association study (GWAS) results are entangled by linkage disequilibrium (LD), and gene expression datasets do not contain genetic variation at individual-level. Here, we propose Multiple Instance Fine-mapping (MIFM) – a multiple instance learning (MIL) objective to overcome the lack of strong labels by grouping putatively causal variants together based on their LD scores. Using MIFM, we trained a deep classifier on a dataset aggregating over 13, 000 GWAS to predict causal variants based on their underlying DNA sequences. We validated variants prioritized by MIFM by constructing polygenic risk scores which transferred better to different target ancestries. Furthermore, we demonstrated how MIFM can be used to disentangle effect sizes of highly-correlated variants to better fine-map GWAS results.

10

POSTER ABSTRACTS

Establishing a system to study landscape of drug-induced somatic mutations in Inflammatory Bowel Disease

Soniya Shende[1, 2,*], Max Brunkhorst[3], Martina Macino[1, 2], Oliver Dyck Dionisi[1, 2], Benedict Monteiro[1, 2], Suharto Banerjee[1], Patrick Weidner[1, 2], Tobias Rolka[1], Carl Weidinger[3], Britta Seigmund[3], and Ashley D. Sanders[4]

1. Max Delbrück Center for Molecular Medicine
2. Berlin Institute of Health
3. Charité - Universitätsmedizin Berlin
4. Max Delbrück Center for Molecular Medicine

* Presenting author

Precision medicine seeks to tailor treatment using genetic and phenotypic insights, but in complex diseases like IBD, tools to predict therapy response remain limited. While germline genetics has been well-studied, recent findings reveal frequent somatic mutations in colonic crypt cells of IBD patients, leading to high clonal diversity. The origins and impact of these mutations on disease progression and treatment response remain unclear. Although factors like aging, environment, and inflammation contribute, it is unknown how IBD therapies themselves shape the somatic genomic landscape. We hypothesize that IBD therapies may drive genomic instability, leading to the accumulation of somatic mutations, clonal selection, and expansion of specific cellular subpopulations. To test this, we developed an in vitro model using patient-derived intestinal organoids (PDOs) from colonic biopsies and a single-cell sequencing method called Strand-seq, which is uniquely sensitive to detecting structural variants (SVs). We first optimized the system by culturing PDOs from healthy individuals for multiple passages and developed methods to characterize the somatic mutations that arise during long-term culture. We analyzed genomic instability per cell and examined mutation types and frequencies to compare subclone composition across passages. This longitudinal study validates our system by revealing mutation types, subclone composition, selection dynamics, potential growth advantages of somatic mutations under culture conditions. We then exposed PDOs to Azathioprine, a commonly prescribed IBD therapy, and observed a notable increase in genomic instability and SVs, suggesting that IBD drugs may themselves contribute to mutagenesis. Going forward, we will use this system to study how additional IBD therapies and inflammatory conditions shape the somatic landscape. Ultimately, this approach integrates high-resolution single-cell genomics with functional patient-derived models to uncover therapy-induced somatic evolution. It offers a novel framework for understanding treatment impact in IBD and other chronic inflammatory diseases, advancing precision medicine through deeper insight into dynamic somatic landscapes.

Divergent Genotype of Hepatitis A Virus in Alpacas, Bolivia, 2019

Talitha Veith[1,*], L. Fabian Beltran-Saavedra[2], Tobias Bleicker[1], Marie Luisa Schmidt[1], José L. Mollericona[2], Kim Grützmacher[3, 4], Rob Wallace[2], Jan Felix Drexler[1, 5], Christian Walzer[6, 7], Terry C. Jones[1, 8], Christian Drosten[1, 5], and Victor M. Corman[1, 5]

1. Institut für Virologie, Charité - Universitätsmedizin
2. Wildlife Conservation Society, La Paz, Bolivia
3. DeutscheGesellschaft für Internationale Zusammenarbeit GmbH
4. Museum für Naturkunde Berlin
5. German Centre for InfectionResearch (DZIF), Berlin
6. Wildlife Conservation Society, New York, USA
7. University of Veterinary Medicine Vienna, Austria
8. University of Cambridge, Cambridge, UK

* Presenting author

Hepatitis A virus (HAV) is a common human pathogen causing acute hepatitis. Human HAV belongs to species Hepatovirus A, genus Hepatovirus, family Picornaviridae. HAV strains of Hepatovirus A are only found in humans (genotypes I-III) and monkeys (genotypes IV-VI), with no non-primate reservoir host. We screened serum and faeces of 70 alpacas and llamas from Bolivia for novel viruses using undirected Illumina High Throughput Sequencing, and detected a divergent HAV. In a PCR screen of 64 alpacas and 6 llamas, we detected HAV RNA in serum and/or faeces of ~9% of alpacas, but not in llamas. In ELISA analysis, we found HAV antibodies in ~64% of alpacas and ~67% of llamas, suggesting HAV infection is common. Complete-genome analysis of alpaca HAV suggests that it is a novel, non-primate genotype of Hepatovirus A. Phylogenetic analysis indicates a long association of HAV with alpacas. HAV in alpacas is similar to human HAV infection: RNA is present in serum and faeces at viral loads comparable to human infection, and immunity builds up over life. Alpaca HAV might belong to the same serotype as genotypes I–VI. Consequently, HAV vaccinations, especially for camelid handlers, should be considered to reduce spillover risk.

Graph Neural Networks in Precision Medicine: Predicting Clinical Outcomes of Infectious Diseases

Natalia von Staa Mansur[1,*], and Helder Nakaya[2, 3]

1. School of Pharmaceutical Sciences – University of São Paulo (USP)
2. University of Sao Paolo
3. Hospital Israelita Albert Einstein

* Presenting author

Understanding how gene expression shapes the immune response to infectious diseases is essential for uncovering mechanisms of pathogenesis, identifying therapeutic targets, and improving diagnostic strategies. Transcriptomic signatures identified through machine learning (ML) techniques are increasingly being explored to support precision medicine. However, the high dimensionality and sparsity of transcriptomic data often lead to overfitting and reduced model stability. It is suggested that considering gene relationships may improve prediction accuracy in biological data. This project aims to explore the use of graph neural networks (GNNs) to model gene expression data within the context of gene-gene relationships, aiming to develop more robust and accurate classifiers for infectious disease outcomes. We focus on Mycobacterium tuberculosis infection, a major global health challenge. Current tuberculosis (TB) diagnostics lack sensitivity and often fail to identify individuals with latent infection who are at high risk of progressing to active disease. RNA-seq data from individuals with varying TB statuses were analyzed, and differentially expressed genes were identified between progressors and non-progressors, as well as between latent and active TB cases. A GNN-based binary classifier is currently being trained to predict clinical outcomes based on both expression profiles and gene network topology. Preliminary results using progressor and non-progressor expression data and an initial PPI matrix reached 95.45% test accuracy, 93.70% validation accuracy, 95.43% test F1-score, and 93.36% validation F1-score. Next steps involve refining the biological network, by integrating selected genes using publicly available protein–protein interaction and co-expression data. Further on, it will be improved model generalization, and applied layer-wise relevance propagation to interpret predictions. The most predictive genes will be functionally analyzed and mapped onto interaction networks to identify key regulatory elements. This approach supports the development of precision diagnostics and deepens our understanding of host-pathogen interactions in TB and other infectious diseases.

Spatial Organization of the Gut Microbiome in Response to Segatella copri Invasion

Michelle Clarissa Cassiopeia Wild1, 2,*, Luiz Gustavo dos Anjos Borges2, Alexander Kukalev3, Sergej Herzog2, Yannik Goerkis1, Johan B. Kantsjö4, Francesca Ronchi4, Ana Pombo3, 5, and Hyun-Dong Chang1

1. German Rheumatology Research Center – A Leibniz Institute (DRFZ), Berlin, Germany
2. Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Berlin, Germany
3. Max-Delbrück Centre for Molecular Medicine, Berlin Institute for Medical Systems Biology, Hannoversche Str. 28, 10115 Berlin, Germany
4. Charité – Universitätsmedizin Berlin, Germany
5. Humboldt-Universität zu Berlin, 10117 Berlin, Germany

* Presenting author

The spatial organization of the gut microbiome plays a critical role in host-microbe interactions and gut health. Current microbiome profiling methods predominantly rely on bulk analyses of stool samples, lacking spatial resolution. Understanding host-microbe interactions within the gastrointestinal tract requires the visualization and quantification of microbial communities in their native spatial context. Dysbiosis, or disruption of this organization, is implicated in various inflammatory and metabolic diseases. Segatella copri (S. copri) is a prevalent gut commensal that has been linked to gut inflammation and systemic immune activation in both humans and mouse models. This study aims to reveal the spatial association patterns in the mouse gut microbiome colonized with S. copri using metaGenomic Architecture Mapping (metaGAM).

Three groups of gnotobiotic Oligo-Mouse-Microbiota (Oligo-MM12) mice, harboring a defined 12-member microbiota community, were analyzed: (1) steady-state Oligo-MM12, (2) Oligo-MM12 colonized with the pathobiont S. copri, and (3) Oligo-MM12 treated with dextran sodium sulphate (DSS). Fecal samples were collected at days 0, 2, and 5 post-DSS treatment and from S. copri-colonized Oligo-MM12 mice for the phenotypic characterization of the gut bacterial community by flow cytometry-based phenotyping. Bacterial cells were stained with an immunoglobulin panel (IgA, IgG, IgM) and a lectin panel (WGA, PNA, STL, ConA) to assess host-microbe and glycan interactions and analyzed by flow cytometry. Fresh-frozen intestinal samples were embedded in OCT, cryosectioned in 10 µm slices, and laser microdissected to generate 384 microsamples. The total DNA was extracted, subjected to whole-genome amplification, and sequenced.

>>> The analysis of the phenotypic data by Bray-Curtis' dissimilarity revealed significant shifts in phenotypic biosignatures over time, indicating dynamic changes in the microbiota during inflammation. MetaGAM analysis revealed the 12 microbial species within defined spatial contexts. Preliminary results demonstrate the utility of spatial microbiome profiling for dissecting the structural and functional complexity of host-microbiome interactions in health and disease.

Using wastewater to study genomic diversity of human viruses

Emanuel Wyler[1,*], Chris Lauber[2,3], Artür Manukyan[1], Aylina Deter[1], Claudia Quedenau[1], Luiz Gustavo Teixeira Alves[1], Claudia Wylezich[4], Tatiana Borodina[1],Stefan Seitz[5,6], Janine Altmüller[1,7], Markus Landthaler[1,8]

1. Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Berlin, Germany
2. Institute for Experimental Virology, TWINCORE Centre for Experimental and Clinical Infection Research, A Joint Venture between the Hannover Medical School (MHH) and the Helmholtz Centre for Infection Research (HZI), Hannover, Germany
3. Cluster of Excellence RESIST (EXC 2155), Hannover Medical School, Hannover, Germany
4. Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Greifswald-Insel Riems, Germany
5. Division of Virus-Associated Carcinogenesis (F170), German Cancer Research Center (DKFZ), Heidelberg, Germany
6. Department of Infectious Diseases, Molecular Virology, University of Heidelberg, Heidelberg, Germany
7. Berlin Institute of Health at Charité, Berlin, Germany
8. Institut für Biologie, Humboldt-Universität zu Berlin, Berlin, Germany

* Presenting author

Viruses infecting humans can not only be found in the human body, but in a range of environmental samples in our surroundings. This includes surfaces such as handrails or door handles, but also indoor air, sewers or the influx of waste water treatment plants. Quantitative detection of viral pathogens in such samples is established since decades. Accessing the wealth of genomic information on human pathogens in environmental samples is however a much more recent development, fueled by the still dropping costs of high-throughput sequencing.

In our previous work, we used total RNA sequencing of longitudinal samplings over 17 months from a wastewater treatment plan in Berlin/Germany in order to elucidate pathogen dynamics over time. Currently, we are focusing on elucidating genomic diversity of enteroviruses and noroviruses in samples from Germany and Bolivia. Here, we present open questions and directions for future research.

Collaborators: Berlin Water authorities, Berlin state office for health and social affairs, noro/entero reference labs at the Robert Koch Institute

MAX DELBRÜCK CENTER BIMSB

Berlin Institute for Medical Systems Biology

Discovery for tomorrow's medicine