
Clinic Meets Data Science
5th Symposium
August 28th 2025
Supported by
PERSONALIZED
HEALTH
BASEL
Sponsored by
University
of Basel
THIS IS
BASEL
BeOne
Roche
SIB
Swiss Institute of
Bioinformatics
biotechnet
switzerland

CMDS

# Clinic Meets Data Science - 5th Symposium

**Date**: August 28th 2025

**Venue**: Oberer Hörsaal, Institute of Pathology, University Hospital Basel, Schönbeinstrasse 40, 4031 Basel, Switzerland

**Time**: 9:00 - 17:30

Dear Participant,

It is our greatest pleasure to welcome you this year in Basel for the 5th edition of the Clinic Meets Data Science symposium.

The event will showcase current developments in clinical data science, highlight state-of-the-art and future applications and tools and discuss possibilities and opportunities that data science offers for the clinical daily routine.

Clinical Data Science focuses on developing computational software, mathematical models, machine learning algorithms and databases for clinical applications. Enhanced digitalization and growing data volumes at hospitals have increased the demand for clinical data science to improve patient care, enhance healthcare quality, and advance clinical research.

The focus of the symposium this year will be on cancer, image analysis, LLMs and data management in the hospital.

This booklet contains logistic information, the detailed program, as well as the abstracts of accepted posters.

The organizing committee

**Organizing committee:**

Gina Faye Boot (University of Basel, University Hospital Basel)
Francesca Faraci (SUPSI-MeDiTech, Lugano)
Abdullah Kahraman (Hochschule für Life Sciences FHNW, Muttenz)
Benjamin Kasenda (University Hospital Basel)
Michal Krauthammer (University of Zürich, University Hospital Zurich)

Contacts: michael.krauthammer@usz.ch
abdullah.kahraman@fhnw.ch

University of Basel

Universitätsspital Basel

Fachhochschule Nordwestschweiz

University of Applied Sciences and Arts of Southern Switzerland
SUPSI

Universität Zürich[UZH]

COMPREHENSIVE CANCER CENTER ZURICH | The University Center for Personalized Oncology in Switzerland

PERSONALIZED HEALTH BASEL

Personalized Health Basel (PHB) is a joint effort of the University of Basel, the University Hospital Basel and the Children Hospital Basel. Its goal is to support local research groups working on personalized medicine and promote translational research. Furthermore, PHB aims to build bridges between academic researchers and private companies to foster collaborations and partnerships and launch pilot-projects in the field of data ecosystem and personalized health.

Attendance of the workshop gives the right to 7 SGMO credit points. Please, request via e-mail the certificate to Abdullah Kahraman of the organizing committee.

# Program

09:00 **Welcome**
Organizing Committee
Basel Tourismus
Cristina Golfieri – Personalized Health Basel

**Session 1: Focus on Cancer**
**Chair: Benjamin Kasenda (USB)**

09:15 **Keynote:**
Judith Zaugg (University of Basel & University Hospital Basel)
"*Systems epigenetics and spatial profiling to study haematological malignancies*"

09:55 **Invited Speaker:**
Maurice Henkel (AI LLM Team, University Hospital Basel)
"*From Data to Dialogue: Implementing Conversational AI for Clinical Decision Support in Oncology*"

10:25 **Selected abstract:**
Gonzalo Cardenal Antolin (ETH Zürich, Switzerland)
"*Large Language Models as Clinical Assistants for Medical Question Answering: A Case Study for HIV*"

10:45 **Coffee break + Poster session**

**Session 2: Focus on Image Analysis**
**Chair: Viktor Koelzer (USB)**

11:15 **Keynote:**
Cristina Granziera (RC2NB, University of Basel & University Hospital Basel)
"*Digital Neurology*"

11:55 **Invited Speaker:**
Andrew Janowczyk (Hôpitaux Universitaires Genève)
"*Implementation of Image Analysis in Clinical Practice*"

12:25 **Selected abstract:**
Jens Wuerfel (Hoffmann-La Roche AG Basel, Switzerland)
"*Federated Learning for Lesion Segmentation in Multiple Sclerosis: A Real-World Multi-Center Clinical Implementation*"

12:45 **Lunch break + Poster session**

# Program

**Session 3: Focus on LLMs**
**Chair: Michael Krauthammer (UZH, USZ) & Abdullah Kahraman (FHNW)**

13:30 **Keynote:**
Michael Moor (ETH Zurich, D-BSSE Basel)
"*Towards grounded medical AI*"

14:10 **Selected abstract:**
Ryan Lusby (Wellcome-Wolfson Institute for Experimental Medicine, Queens University, Belfast, United Kingdom)
"*CNN-based learning of single-cell transcriptomes reveals a blood-detectable multi-cancer signature of brain metastasis*"

14:40 **Coffee break + Poster session**

**Session 4: Focus on Translation into the Clinic**
**Chairs: Bram Stieltjes (USB)**

15:10 **Keynote:**
Lars Olsen (University of Copenhagen)
"*From innovation to implementation: bridging data science and clinical cancer care in Denmark*"

15:50 **Invited Speaker:**
Waldo Valenzuela (Data Science Center, InselSpital University Hospital Bern)
"*Medical-Blocks: A platform for Translational Research and Clinical Data Science*"

16:20 **Invited Speaker:**
Joshy Cyriac (Department of Theragnostics, University Hospital Basel)
"*Challenges in the Clinic- a personal retrospective on clinical software engineering*"

16:50 **Panel Discussion:** Future Challenges of Data Science in Clinics
Chairs: Francesca Faraci, Gina Faye Boot, Beatrice Zanchi

17:30 **Closing remarks & Apero**

# Abstracts of accepted posters

# Abstracts

## Bridging Experimental Research and Clinical Decision-Making: An Integrated IT Solution for Translational Medicine

**Shan Yang** (1,3), Yannick Blum (2), Andrea Mock Caceres (2), Maren Diepenbruck (2), Loïc Sauteur (2), Raffaello Ferone (3), Cristina Golfieri (3), Mohamed Bentires-Alj (2), Bram Stieltjes (1,3), Charly André Jehanno (2), Benjamin Kasenda (4)

1Department of Digitalisation and ICT, University Hostpital Basel
2Department of Biomedicine, University and University Hospital Basel
3Personalized Health Basel
4Department of Medical Oncology, University and University Hospital Basel

**Background and Challenge**
Translational medicine seeks to rapidly convert laboratory discoveries into clinical applications that benefit patients. However, a persistent gap between novel findings and their routine use in patient care complicates this mission. To address it, we developed an IT solution that bridges experimental data from research laboratories and the clinical teams who rely on it for evidence-based decisions in early-phase trials.

**Solution and Core Novelty**
The pipeline integrates patient-derived tumor organoid (PDTO) drug-sensitivity assays into routine care by leveraging two proven institutional platforms—REDCap and OMERO. REDCap serves as the authoritative repository for patient identifiers, clinical metadata and longitudinal study records, whereas OMERO securely manages microscopy images and scan-level metadata. An API-driven exchange synchronises identifiers, assay results and image references, linking research output with strict patient-data safeguards while eliminating extra databases and manual transfers. In addition, dose-response metrics (IC50, AUC) are automatically summarised and pushed to electronic case-report forms.

**Security and Data Integrity**
The system follows institutional information-security policies: all sensitive content resides on protected university-hospital servers, preventing unauthorised egress. Seamless integration avoids new data silos and maintains audit-ready, regulator-approved provenance—each reported result can be traced unambiguously to its original image, assay plate and patient record.

**Impact and Future Implications**
By streamlining the data flow from laboratory bench to clinical trial application, the solution significantly accelerates the translation of experimental insights into real-time clinical decision-making. The resulting framework provides a reusable blueprint for future translational technologies, enabling institutions to transform laboratory innovation into reliable, auditable clinical routine.

# Abstracts

## AI-Driven Detection of Distinct Disability Patterns Within Equivalent EDSS Scores

**Martina Greselin** (1,2,3), Po-Jui Lu (1,2,3), Magdalena Mroczek (2,3), Nuria Cerdá-Fuertes (1,2,3), Anastasios Demirtzoglou (2,3), Athina Papadopoulou (1,2,3), Jens Kuhle (2,3), David Leppert (3), Sophie Arnould (4), Manar Aoun (4), Ludwig Kappos (2,3), Cristina Granziera (1,2,3), Marcus D'Souza (2,3)

1Translational Imaging in Neurology (ThINk) Basel, Department of Biomedical Engineering, Faculty of Medicine, University Hospital Basel and University of Basel, Switzerland.
2Department of Neurology, University Hospital Basel, Switzerland.
3Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), University Hospital Basel and University of Basel, Switzerland.
4Novartis Pharma AG, Basel, Switzerland.

**Background:**
The Neurostatus-Expanded Disability Status Scale (EDSS) is the most widely used tool to measure disability in multiple sclerosis (MS) clinical trials. However, EDSS scores of 4.5 or higher are heavily influenced by ambulation ability and may not adequately reflect impairments in other functional domains.

**Objective:**
To investigate whether assessments assigned the same EDSS score can represent distinct patterns of disability.

**Methods:**
We analyzed 13,103 clinical assessments from 1,636 individuals with secondary progressive MS enrolled in the EXPAND trial, a double-blinded, randomized, phase 3 study, where 1651 people with secondary progressive MS (pwSPMS) were randomly assigned to receive Siponimod or placebo. Each assessment included Functional System Scores (FSS), their respective subscores, Ambulation Scores, and EDSS scores. We first performed a descriptive analysis to identify the most relevant Functional Systems. Subscores were then binarized following Neurostatus criteria and grouped by EDSS score. A two-step machine learning clustering approach was applied to identify patterns. Clusters were subsequently aggregated based on dominant subscore features to define new subscore patterns.

**Results:**
Clustering revealed multiple assessment groupings with similar subscore profiles. Among individuals with EDSS scores of 4.0 and above, four distinct subscore patterns emerged within the same EDSS score range.

**Conclusion:**
Leveraging artificial intelligence to analyze large-scale, high-quality clinical data enables the identification of nuanced subscore patterns within identical EDSS scores, offering a more detailed view of disability in MS.

# Abstracts

**Machine Learning-Based Segmentation of Adipocytes in Imaging Data**

**P. Kos**, J.Haslbauer, J.Zaugg

Department of Biomedicine, University of Basel, University Hospital Basel, 4031 Basel, Switzerland

Aberrant signaling across diverse cell types and states contributes to the progression of many diseases. In acute myeloid leukemia (AML), adipocytes remain underexplored despite growing evidence linking obesity to poor clinical outcomes [1, 2]. Although traditionally viewed as passive fat-storing cells, adipocytes may actively influence the tumor microenvironment. The underlying molecular mechanisms remain unclear.

In imaging datasets from AML patients, most studies focus on other cell populations. Standard segmentation tools such as StarDist or Cellpose, rely on nuclear features [3], which makes them suboptimal for detecting adipocytes. They often have displaced or indistinct nuclei. As a result, adipocytes are frequently overlooked or inaccurately segmented.

Here, we present a machine learning approach specifically designed to identify adipocytes in histological images. We manually annotated whole-slide H&E images to train a model capable of generalizing across imaging modalities, including immunofluorescence. On an independent validation set, the model achieved a mean Intersection over Union (IoU) of 95%, demonstrating high segmentation accuracy. This approach enables more accurate adipocyte identification and facilitates investigation into their role in AML pathophysiology.

References

1. Hopkins BD, Goncalves MD, Cantley LC. Obesity and cancer mechanisms: cancer metabolism. J Clin Oncol. 2016;34(35):4277–4283.
2. Zhang Z, et al. Bone marrow adipose tissue-derived stem cell factor mediates metabolic regulation of hematopoiesis. Haematologica. 2019;104(9):1731.
3. Schmidt U, et al. Cell detection with star-convex polygons. In: MICCAI 2018, Springer; 2018:265–273.

# Abstracts

**Oncology related conceptual grounding for large language models**

**Aman Sinha**, Bogdan-Valentin Popescu, Xavier Coubez, Marianne Clausel, Mathieu Constant
University of Lorraine, France

Language models (LMs) capabilities have grown with a fast pace over the past decade leading researchers in various disciplines, such as biomedical research, to increasingly explore the utility of LMs in their day-to-day applications. Domain specific language models have already been in use for biomedical natural language processing (NLP) applications. Recently however, the interest has grown towards medical language models and their understanding capabilities. In this study, we investigate the medical conceptual grounding of various language models against expert clinicians for identification of hallmarks of immunotherapy in breast cancer abstracts. Our results show that pre-trained language models have potential to outperform large language models in identifying very specific (low-level) concepts.

# Abstracts

**Resolving Tumor Clonality in NSCLC through Whole-Exome Sequencing**

**Luca Roma**, Obinna Chijioke, Spasenija Savic Prince, Lukas Bubendorf
Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, Switzerland.

**Introduction:**
Tracing clonal trajectories remains a major challenge in non-small cell lung cancer (NSCLC), especially in histologically heterogeneous subtypes such as lung adenocarcinoma (LUAD), pulmonary pleomorphic carcinoma (PPC), adenosquamous carcinoma (ASC) and in patients presenting with multiple synchronous or metachronous tumors.

**Methods:**
We applied whole-exome sequencing (WES) to investigate the clonal relationship of rare NSCLC cases, including PPC and ASC with biphasic histology, and a longitudinal case of a patient with three metachronous EGFR-mutant LUAD. Somatic mutations and copy number alterations were analyzed to assess clonal relationships. Retrospective histopathological review was conducted to validate molecular findings.

**Results:**
In the EGFR-mutant LUAD case, all three tumors harbored the EGFR L858R mutation but showed divergent mutational and copy number profiles, indicating independent clonal evolution despite a shared therapeutic target. This might have a potential impact on treatment and disease monitoring. In PPC and ASC, distinct histological components shared multiple truncal mutations in key driver genes, supporting a monoclonal origin. Molecular findings were consistent with histopathological assessments.

**Conclusions:**
This study highlights the power of WES to resolve tumor clonality in NSCLC. The EGFR-mutant LUAD case emphasizes that a single shared EGFR driver mutation alone is not enough to demonstrate the shared origin, with direct implications for clinical management. Furthermore, clonality analysis improves our understanding of tumor heterogeneity in aggressive subtypes such as PPC and ASC.

# Abstracts

**Empowering Clinical Research with LLMs: Good Practices in AI-assisted Data Analysis**

**Jelena Čuklina***, Peter Krusche*

*Novartis Pharma AG

Large language models (LLMs) have great potential to transforming clinical data analysis – they bring sophisticated capabilities to help translating scientific questions into concrete program code that interrogates clinical datasets and outputs answers. In this presentation, we share practical insights into how LLMs can enhance data science workflows — from research problem scoping to code generation, troubleshooting, and reproducibility. We will demonstrate how AI tools can help refine research questions, reveal logical gaps and quickly set up structured and well documented code repositories.

One pitfall we would like to discuss is around computational infrastructure specificity, and how to use software package dependencies from a large and evolving ecosystem: LLMs typically have seen a particular version of their documentation during training which may not be appropriate to the environment a user works in.

To address this, we have developed approaches that can provide environment-specific context to AI assistants. We focus on the R programming language and extract R package documentation for augmenting the conversation context and prompt engineering. This improves the accuracy of AI-generated output. Additionally, we discuss AI assistance in fosteringgood collaborative practices and analytical rigor – for example, lowering the barriers to use appropriate tools that aid reproducibility.

Our experience shows that thoughtful integration of AI into clinical workflows enhances not only efficiency but can also embed deeper analytical thinking and assist with analysis reproducibility. AI coding assistants are not merely accelerators—they are catalysts for better data science.

# Abstracts

**Single-cell transcriptomic analysis of the bone marrow stromal networks during chemotherapy**

Ana Luísa Pereira (1), Ute Süssbier (1), **Karolina Zielinska** (1), Anjali Vijaykumar (1), Paul Büschl (1,2), Alvaro Gomariz (1,2), Thomas Zerjatke (3), Patrick Helbling (1), Stephan Isringhausen (1), Hui Chyn Wong (1), Takashi Nagasawa (4), Yokomizo Tomomasa (5), César Nombela-Arrieta (1)

1 Department of Medical Oncology and Hematology, University Hospital and University of Zürich, Zürich, Switzerland
2 Computer Vision Laboratory, ETH Zurich, Zurich, Switzerland
3 Faculty of Medicine Carl Gustav Carus, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany
4 Laboratory of Stem Cell Biology and Developmental Immunology, Graduate School of Frontier Biosciences and Graduate School of Medicine, Osaka University, Osaka, Japan
5 International Research Center for Medical Sciences, Kumamoto University, Kumamoto, Japan

One of the earliest drugs administered to patients with cancer was 5-fluorouracil (5-FU). Today 5-FU is still used to treat cancer and in preparation for bone marrow transplantation. 5-FU eliminates highly proliferative hematopoietic progenitor cells triggering a reactive regenerative response from hematopoietic stem cells, which eventually restores hematopoietic function. While the responses of the hematopoietic compartment have been studied in detail, the impact of chemotherapy on specific stromal components of the bone marrow (BM) microenvironment is less understood. Here, we employ customized 3D microscopy and scRNA-seq to study the effects on endothelial cells (ECs) and
mesenchymal stomal cells (MSCs) and analyze their regenerative dynamics.
We performed scRNAseq analysis of stromal cells isolated at 14, 56 and 120 days post challenge. At all time points we could detect eight MSC subsets based on previously described signatures, of which the so called adipo-Cxcl12-abundant reticular cells (adipoCARc) were the most represented. A detailed analysis of adipoCARc revealed three clusters corresponding to different cell states and distinct time point composition while the diffusion map algorithm established a trajectory. Pathway analysis showed
the activation of inflammatory pathways at the starting point of the trajectory after challenge with 5-FU. Next, we aimed to identify the temporal expression patterns of genes involved in the regenerative response. We detected genes following three trends: peak, recession and growth with recession being the most common trend. The ORA analysis of recession genes identified enrichment in inflammatory signaling.
We next investigated the dynamics of the endothelial compartment focusing on ECs from sinusoidal vessels (SECs). We identified three clusters of SECs with different composition of time points post 5-FU treatment. Trajectory analysis revealed two cell states corresponding to recovery and one state corresponding to acute injury. Subsequent pathway analysis along the trajectory identified epithelial to mesenchymal transition and Myc signaling as the top pathways that change their activity. We also investigated the long term effects of 5-FU and found that in both SECs and adipoCARc transcriptome remains long-affected post treatment. Our findings describe acute and long-term transcriptomic changes in stromal cells that uncover a tissue regenerative program leading to the reactivation of BM
hematopoietic function.

# Abstracts

**Integration of Health Data Through Scalable and Reproducible Analysis Pipelines**

**Antoine Buetti-Dinh** (Zurich Data Scientists)
Klaus Steigmiller (Zurich Data Scientists)
Matteo Tanadini (Zurich Data Scientists)

Health data, including genomics and clinical phenotype information, has become increasingly critical in driving advancements in medical research and personalized healthcare. However, efficiently processing and integrating such complex datasets requires scalable and reproducible computational workflows. Here we present the development of analysis pipelines that address these challenges by leveraging cutting-edge tools like Nextflow, nf-core, and Seqera Platform. Such pipelines rely on containerization technologies to create a flexible framework for biomedical data processing applicable to a wide range of computational environments.

# Abstracts

**VAEriety: Biologically-informed generative deep learning for clinical translatability of animal models**

**Jonas Meirer** (CSEM), Bark Sahin (CSEM), Lucas D. Wittwer (CSEM), Tim Heinemann (CSEM), Sohyon Lee (KAIST), Carine Poussin (CSEM)

Animal models are essential to biomedical research, offering insights into disease biology for therapeutic development. However, biological differences between species often limit how well findings in mice translate to humans. This challenge is especially pronounced in glioblastoma, a rare and heterogeneous brain tumor, where progress has been slowed by the scarcity of human samples and the limitations of existing models to mirror human disease dynamics.

As part of the VAEriety project, we are developing a machine learning framework to bridge this gap by computationally translating transcriptomic profiles from mouse models into their human counterparts. Our approach is enabled by large-scale single-cell RNAseq databases such as CELLxGENE, including millions of cells from both species. To translate disease characteristics across species, we are designing a conditional neural network that integrates the structure of known biological pathways from the REACTOME database into its architecture for improved robustness and interpretability. The conditional setup further allows us to understand relevant biology in a species-dependent way, while discarding irrelevant variation such as batch effects.

The design improves interpretability of the model's outputs, enforces learning biologically plausible disease characteristics, and allows assessment of model quality through pathway-level comparisons with known biology. By creating a tool to translate transcriptomic signals across species, we aim to uncover conserved molecular disease mechanisms in glioblastoma and pinpoint features unique to each organism. Ultimately, this work aims to strengthen the translational relevance of animal studies and support the identification of therapeutic targets with greater relevance to human disease.

# Abstracts

**CnQuant: A lightweight stack of 5 applications to streamline the analysis of Illumina Infinium Methylation array data in oncological practice**

**Freyter BM**, Hultschig C, Brugger J, Frank S, and Hench J. Institute for Medical Genetics and Pathology at University Hospital Basel

DNA methylation and chromosomal copy number (CN) profiling are standard for personalized cancer diagnostics, particularly for central nervous system tumors. Illumina Infinium Methylation bead-based arrays (450k, EPICv1, EPICv2) enable quasi genome-wide methylation analysis, with CN variations inferred using CN-neutral reference samples. Existing CN variation-calling tools demand computational expertise, limiting accessibility. CnQuant, our scalable, user-friendly software stack, addresses these challenges for clinical use. It comprises five modular applications: CQcalc unifies 450k, EPICv1, and EPICv2 array formats, infers copy number variation (CNV) and generates single-case plots, viewable interactively at single-probe resolution via CQcase. CQall_plotter detects recurrent CNVs within tumor entities, visualized through CQall. CQcase and CQall enable view sharing via URL or exported interactive HTML plots, annotation of individual genes, selecting bins or probes, and provides tabular data access to GeneCards links. CQmanager seamlessly integrates into clinical workflows by accepting POST requests from custom scripts, enabling CNV analysis automation according to user-specific needs. CQmanager features a queuing system in-parallel running CQcase containers, adaptable to the available hardware capabilities. Available as a free web service, Docker container, WSL image, and open-source code, CnQuant enhances automated CNV inference, genomic annotation, and cohort comparison for routine diagnostics. CN profiling might reveal oncologically relevant alterations guiding therapy, facilitating tumor classification in doubtful situations, and can guide in-depth DNA/RNA targeted sequencing for selected patients. We consider CnQuant to be of significant practical value for the oncological community.

# Abstracts

**Estimating the Charlson Comorbidity Index (CCI) from Diagnoses in Real-world EHR Data using LLM**

**Michael Rebsamen**, Lana Borcard, Benedikt Herzog, Anna-Katharina Calek, Mazda Farshad, David Bauer, Sebastiano Caprara [Balgrist University Hospital]

**Background and Motivation:**
CCI can be a metric of interest for cohort selection in research projects. However, the CCI is usually not available for real-world data in EHRs, and manual calculation is not viable. We assessed the feasibility of leveraging large language models (LLMs) to automatically derive the CCI from patient diagnosis lists extracted from EHR systems.

**Materials and Methods:**
List of diagnoses and age groups were extracted from the clinical EHR for a total of 200 patients, randomly selected from each ASA class I-IV. Three MDs manually rated all cases, followed by consensus among the raters. Diagnoses and age group served as input to various LLMs along with instructions to estimate the CCI.

**Results:**
Preliminary results on the validation dataset revealed excellent inter-rater agreement among human experts (ICC = 0.97). The accuracy of LLMs generally improved when specific guidelines for each category were provided to calculate the CCI. The highest scores were achieved with Llama 4 Maverick ($R^2 = 0.93$) and GPT-4o ($R^2 = 0.89$).

**Conclusions:**
LLMs require carefully crafted prompts with guidelines to calculate the CCI reliably. The results suggest that the CCI can be derived automatically at large scale for research purposes.

# Abstracts

**Exploring the potential of USB-derived synthetic clinical data**

**Maren Diepenbruck*** (1), Marc Bovet* (1), Fabian Franzeck* (1), Alexander Kleer (1), Andreas Aeberhard (1), Dana Yaffe (2), Yuri Gavrielov (2), Leenor Alfahel (2), Ola Yakov (2), Nadia Vigdar (2), Yulia Haidenraich (2), Raffaello Ferone (3), Cristina Golfieri (3), Bram Stieltjes (1,3)

*Equal contribution
1 University Hospital Basel (USB), Basel, Switzerland
2 MDClone Ltd., Beer Sheva, Israel
3 Personalized Health Basel, University Basel, Basel, Switzerland

The University Hospital Basel (USB), within the framework of the Swiss Personalized Health Network (SPHN), has developed a state-of-the-art clinical data warehouse (CDWH) that aggregates structured health data from more than 100 primary systems, facilitating operational improvements, diverse research and innovation efforts.

Over the past years, a partnership between USB and MDClone, an Israeli company specializing in privacy-preserving synthetic data generation, has culminated to enhance USB's clinical data usage via MDClone's innovative ADAMS platform and synthetic data engine. This platform allows partners to independently explore, filter, and extract statistically representative synthetic data derived from USB's CDWH. A recently completed pilot project with Roche validated the platform's utility, and an ongoing internal proof-of-concept study continue to explore its application internally. The synthetic data approach ensures robust patient privacy, maintaining HIPAA/GDPR compliance and facilitates rapid access to clinical data and supports secondary data usage to accelerate translational research, algorithm development and clinical innovation. Moreover, the infrastructure is well-suited to enable federated learning and federated querying across institutions, thus fostering collaborative, privacy-aware data science initiatives.
In conclusion, USB-derived synthetic clinical data promote scalable and responsible clinical data usage and the overall vision towards a data-driven hospital and a democratization of the health ecosystem.

# Abstracts

**ACTIBATE randomized controlled trial: Personalized response prediction to a 24-week exercise intervention in young healthy adults using multiomics**

**Gavin Graf** [1], Falko Noé [2,3], Vincent Gardeux [4], Alaa Othman [3], Bart Deplancke [4], Jonatan R. Ruiz [5], Christian Wolfrum [2], Ferdinand von Meyenn [1], Adhideb Ghosh [1,2]

1. Laboratory of Nutrition and Metabolic Epigenetics, Institute of Food, Nutrition and Health, Department of Health Sciences and Technology, ETH Zurich, Switzerland
2. Laboratory of Translational Nutrition Biology, Institute of Food, Nutrition and Health, Department of Health Sciences and Technology, ETH Zurich, Switzerland
3. Functional Genomics Center Zurich, University of Zurich and ETH Zurich, Switzerland
4. Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, EPFL, Switzerland
5. PROFITH (PROmoting FITness and Health through Physical Activity) Research Group, Department of Physical Education and Sport, Faculty of Sport Sciences, University of Granada, Spain

Regular physical activity is a key component of preventive health, known to reduce mortality and the risk of chronic diseases by improving cardiovascular, metabolic and immune function. While aerobic and resistance training enhance body composition and metabolic health, individuals show highly variable responses to exercise, even in the absence of overt disease. This variability underscores the need for personalized interventions to optimize health outcomes.

To investigate the biological basis of this variability, we analyzed data from the ACTIBATE study, a randomized controlled trial involving approximately 100 sedentary young adults (18–25 years, 50% women) who completed a 24-week supervised aerobic and resistance training program at moderate or vigorous intensity. Comprehensive clinical assessments, plasma samples, and biopsies of subcutaneous adipose tissue and skeletal muscle were collected at baseline and post-intervention. Although exercise did not significantly alter brown adipose tissue (BAT) activity, it led to reductions in body weight, fat mass, and visceral adipose tissue, independent of exercise intensity.

To explore individual differences in response, Multi-Omics Factor Analysis (MOFA) was applied to baseline bulk transcriptomic, metabolomic, and lipidomic datasets. This unsupervised approach revealed distinct participant subgroups, primarily differentiated by LDL cholesterol levels and BMI, with meaningful differences in baseline metabolic profiles, clinical outcomes, and DNA methylation. These findings demonstrate the potential of multi-omics profiling to identify predictive biomarkers and stratify individuals prior to intervention, supporting more personalized and effective exercise strategies for improving metabolic health.

# Abstracts

**CNN-based learning of single-cell transcriptomes reveals a blood-detectable multi-cancer signature of brain metastasis (**Selected for an oral presentation)

**Ryan Lusby** (1), Sarah Carl (2), Vijay K. Tiwari (1,3,4,5,6)

1. Wellcome-Wolfson Institute for Experimental Medicine, School of Medicine, Dentistry & Biomedical Science, Queens University, Belfast, BT9 7BL, United Kingdom.
2. Scailyte AG, Lichtstrasse 35, 4056 Basel, Switzerland
3. Institute of Molecular Medicine, University of Southern Denmark, Odense C, Denmark
4. Patrick G Johnston Centre for Cancer Research, Queen's University, Belfast BT9 7AE, UK.
5. Danish Institute for Advanced Study (DIAS), Odense M, Denmark
6. Department of Clinical Genetics, Odense University Hospital, Odense C, Denmark

Brain metastasis (BrM) is a serious complication of advanced cancers and remains difficult to predict before clinical symptoms appear. To investigate shared transcriptional features of BrM across tumour types, we integrated single-cell RNA sequencing (scRNA-seq) data from malignant epithelial cells derived from six carcinoma types, including lung, breast, colorectal, renal, prostate, and melanoma. We applied ScaiVision, a supervised deep learning method, to classify tumour cells based on BrM status. The models achieved high predictive accuracy (area under the ROC curve > 0.90) across all six cancer types. This analysis identified a consistent pan-cancer gene expression signature associated with BrM, defined at single-cell resolution. To evaluate the clinical relevance of this signature, we assessed its presence in tumour-educated platelets (TEPs) from blood samples of patients with and without BrM. The signature was detectable in platelet RNA and distinguished patients with BrM from those without, indicating that features of the BrM-associated expression program are reflected in blood-derived material. These findings demonstrate that a transcriptional signature of brain metastasis can be identified across multiple tumour types using scRNA-seq and neural network-based analysis. The detectability of this signature in TEPs supports its relevance in a non-invasive context and provides a basis for further investigation into its utility for BrM risk assessment.

# Abstracts

**Large Language Models as Clinical Assistants for Medical Question Answering: A Case Study for HIV**
(Selected for an oral presentation)

**Gonzalo Cardenal Antolin** (1), Jacques Fellay (2,3,4), Bashkim Jaha (5,6), Roger Kouyos (5,6)*, Niko Beerenwinkel (1,3)*, Diane Duroux (7,1,8,3)*

1 Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland;
2 School of Life Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland;
3 Swiss Institute of Bioinformatics, Lausanne, Switzerland Biomedical Data Science Center, Lausanne University Hospital;
4 University of Lausanne, Lausanne, Switzerland;
5 Institute of Medical Virology, University of Zurich, Zurich, Switzerland;
6 Department of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland;
7 ETH AI Center, ETH Zurich, Zurich, Switzerland;
8 Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland;
* Equal contribution

Large language models (LLMs) offer promising support for clinical decision-making, particularly in complex areas like HIV management, which involves numerous treatment options, comorbidities, and adherence challenges. However, their integration into practice is hindered by concerns around accuracy, potential harm, and clinician trust. Despite growing interest, the performance of LLMs in HIV care remains underexplored, with limited benchmarking.

To address this, we developed HIVMedQA, a benchmark for evaluating LLMs on open-ended questions related to HIV patient management. The dataset was curated and validated by an infectious disease physician. We assessed seven general-purpose and three medically specialized LLMs using tailored prompt engineering and multiple evaluation metrics, including both lexical similarity and LLM-as-a-judge scoring. We focused on key aspects such as comprehension, reasoning, factual accuracy, bias, and potential harm. Our results show that Gemini 2.5 Pro outperformed all other models across most evaluation criteria. Interestingly, medically fine-tuned models did not always surpass general-purpose ones, and larger model size did not guarantee better performance. Reasoning and comprehension were more challenging than knowledge recall, and models exhibited cognitive biases like recency and status quo effects. Using an LLM as a judge provided more clinically relevant assessments than traditional scoring methods.

These findings highlight the limitations of current models and the need for targeted improvements in LLM development and evaluation to ensure safe, effective clinical integration.

# Abstracts

**Federated Learning for Lesion Segmentation in Multiple Sclerosis: A Real-World Multi-Center Clinical Implementation (**Selected for an oral presentation)

**Jens Wuerfel** - F. Hoffmann-La Roche AG, Basel, Switzerland

Sarah Hindawi - Hoffmann-La Roche Limited, Mississauga, ON, Canada

Bartlomiej Szubstarski - Roche Polska Sp. z o.o., Warsaw, Poland

Bjoern Tackenberg, Jens Wuerfel, Eric Boernert - F. Hoffmann-La Roche AG, Basel, Switzerland

Multiple sclerosis (MS) is a chronic neurodegenerative disease driven by immune-mediated central nervous system damage, often leading to progressive disability. Accurate segmentation of MS lesions on MRI is crucial for disease monitoring and treatment efficacy; however, manual segmentation remains time-consuming and prone to variability. While deep learning has advanced automated segmentation, robust performance benefits from large-scale, diverse datasets, yet data pooling is restricted by privacy regulations and clinical performance remains challenged by inter-site heterogeneity. This real-world study evaluated a federated learning (FL) framework for MS lesion segmentation, using the self-configuring nnU-Net model collaboratively trained on 380 MRI cases from three institutions without sharing raw patient data. The federated model, trained for 50 rounds, achieved Dice scores ranging from 0.66 to 0.80 across held-out test sets. While performance varied across sites, reflecting data heterogeneity, the study demonstrates the potential of FL as a scalable and secure paradigm for advancing automated MS analysis in distributed clinical environments. This work supports adopting secure, collaborative AI in neuroimaging, offering utility for privacy-sensitive clinical research and a scalable template for medical AI development, bridging the gap between model generalizability and regulatory compliance.