INTERNATIONAL EDITION

gesundhyte.de

THE MAGAZINE FOR DIGITAL HEALTH IN GERMANY

ISSUE 15 DECEMBER 2023

focus: research reconnected!

from page 12

preventing common disease with metabolomics and AI

page 44

with system against addictive disorders

page 68

scars on the way – fathoming fibrosis with AI and omics

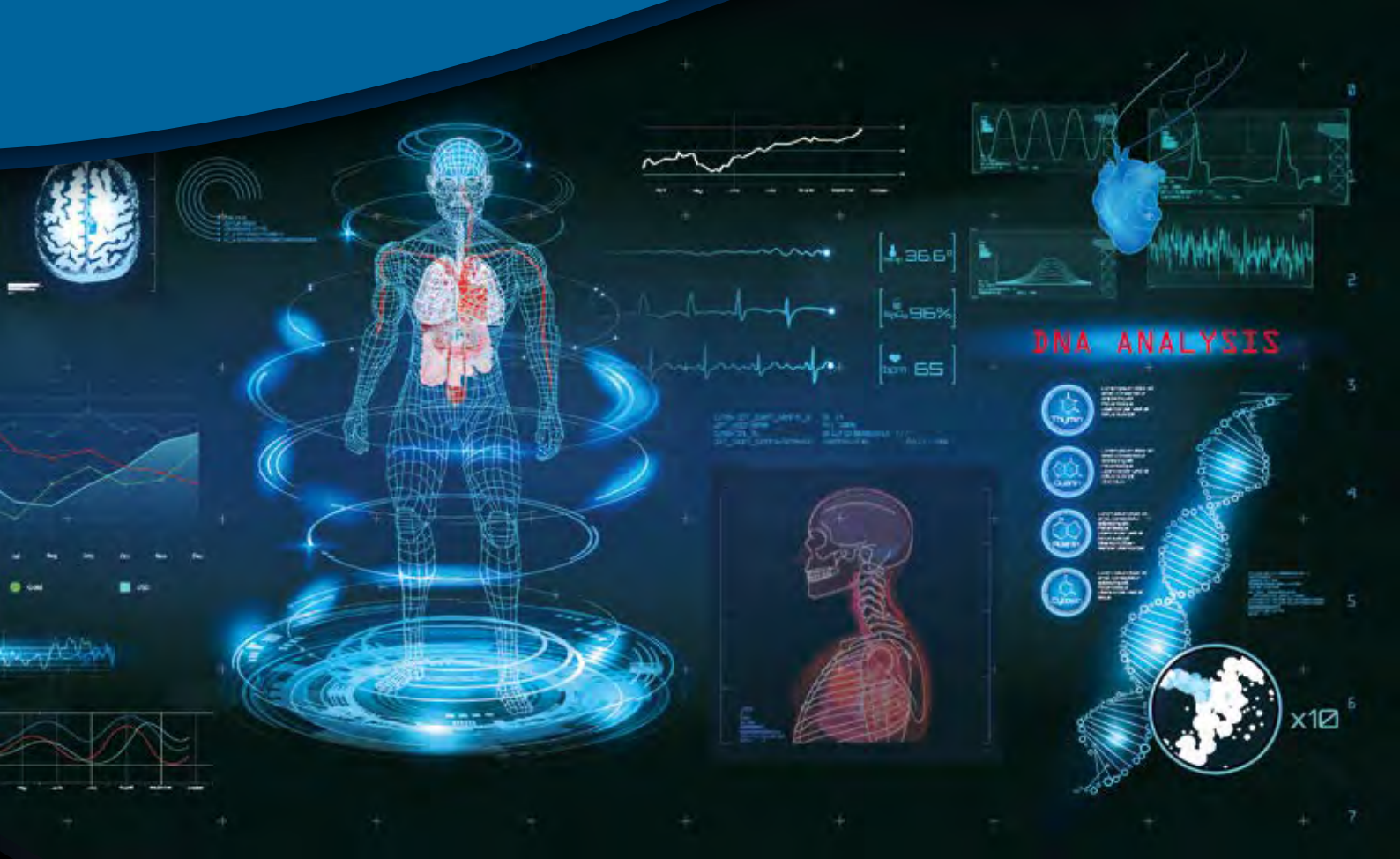
page 76

interviews with

Beatrice Lugger
Katarzyna Bozek
Maike Sander
Oliver Stegle

page 8, 40, 64 and 92

ISSN 2748-3525



gesundhyte.de

is dedicated to finding solutions for today's most important challenges in health care focusing on Digital Health and Systems Medicine, two young disciplines that are considered to form the medicine of the future. The key is the combination of laboratory research data of different kinds with real-world data – from bench to bed-site. Innovative technologies and methods will pave the way for more precise predictions and personalised therapy. Already today, this approach is successfully used in Oncology and will be extended to other diseases in the future. Here, the establishment of robust and standardized IT infrastructure plays a major role to allow for the secure exchange of patient data with research teams. Read the magazine gesundhyte.de to find out how this innovative branch of science works to provide solutions for our current and future challenges in medicine.

Cover photo: AdobeStock © amiak

Photo: AdobeStock © SergeyBitos

greetings

Dear Readers,

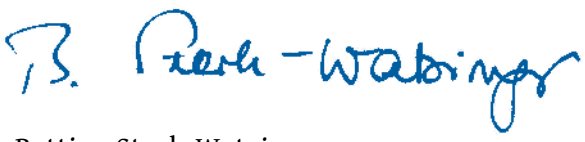
The insight that the whole is greater than the sum of the parts applies equally to the development of research landscapes. Cross-disciplinary networks help pool efforts, achieve synergies and produce new solutions. Networking injects momentum into medical progress so that people can enjoy the fruits of medical research more quickly. It is precisely this approach that the funding measures of the Federal Ministry of Education and Research have in common.

The focus of this edition of gesundyte.de – **"Research reconnected"** – applies especially to the digital transformation in medicine. A key role in this is played by our Medical Informatics Initiative which provides enormous impetus for digitalization in health research. At the start of the year it entered its consolidation and extension phase with the clear goal of even greater networking. The intention is for the initiative to drive the development of a national and decentralized infrastructure for health research data through close cooperation with other pioneering initiatives, especially the national Network of University Medicine (NUM).

In order that patients will be able to experience the benefits of innovations not only at university hospitals but also more widely, the "Digital Hubs: Advances in Research and Health Care" are developing pilot models for the transfer of innovations to local outpatient care. This too can only work where there is good networking of all the regional health care partners, from university hospitals to family doctors and rehabilitation care.

Interdisciplinary exchange is at the heart of many more measures supported by my Ministry. One example is the Modeling Network for Severe Infectious Diseases. The SARS-CoV-2 pandemic showed very clearly how important modeling is for gaining new understanding as early as possible of the development of infection rates and the effectiveness of countermeasures. It requires the collaboration of experts from medicine, mathematics and informatics.

This issue of gesundhyte.de provides you with a wide range of news and articles on how networked research speeds up medical progress with the help of modern information technology. Let yourself be inspired by what is already possible now and what will be possible in the near future.



Bettina Stark-Watzinger
Member of the German Bundestag
Federal Minister of Education and Research

Photo: Bundesregierung – Guido Bergmann

greetings

Dear Readers,

Even in 2023, the aftermath of the Corona pandemic is still clearly felt within the healthcare system. Nevertheless, there is reason to look positively toward the future: after all, we will take a big step closer to our mission of strengthening the healthcare system if we mobilize previously dormant potential.

One example: In Germany's largest university hospital, the Charité, significant amounts of routine data are collected daily during day-to-day operations. So far, this data has mostly been used only selectively and not systematically enough for research and improving care. Unfortunately, this is not an isolated case and applies to many other hospitals in Germany. Systematically tapping into this treasure trove of data would be an enormous gain. Therefore, the expansion of a network for the sensible, responsible and targeted use of collected routine data is essential. The first steps have already been taken. The Medical Informatics Initiative of the German Federal Ministry of Education and Research was established in 2018 with the purpose of making the best possible use of digitization in medicine for care and research. To achieve this goal, data integration centers have been established at university hospitals and partner institutions, which are interconnected with each other. With the continuation of the Network of University Medicine and the integration/fusion of the data integration centers from the Medical Informatics Initiative, clinical networking is now secured in the long term.

Another funding measure from the Medical Informatics Initiative, in which the Charité plays a major role, provides significant support for networking within the clinics: the HiGHmed consortium aims to find novel, interoperable solutions in medical informatics, with the goal of making medical patient data accessible for clinical research and teaching. This initiative benefits better and future-oriented patient care. The aim is to advance digitization in Germany without losing sight of the well-being of patients. The data is already available in the clinical context; there's no need to collect it again. We now have to ensure that it is used in research in a targeted and safe manner. Charité's Strategy 2030 – *"Rethinking Health"*, aims to do just that: *"We are overcoming disciplinary, conceptual and structural boundaries to improve healthcare. In doing so, we combine science and caring [...] We aim for robust and relevant research results and are committed to trust in scientific knowledge."*

Discover more about advancements in digitization and data networking in this interesting new issue of gesundhyte.de.

I wish you a stimulating read!

Prof. Dr. Heyo K. Kroemer
Chief Executive Officer, Charité – Universitätsmedizin Berlin

Photo: Wiebke Peitz / Charité

foreword

Dear Readers,

Either we find a way, or we make one. What sounds like a motto of a German politician is actually a saying of the Carthaginian general Hannibal Barkas (around 247-183 BC). Somewhat more peacefully, we could translate this saying with the bon mot, *"where there is a will, there is a way"*. This way the present issue of gesundhyte.de is also to be understood. Crosslinking research in a new way means, with the good will of all involved, to further develop proven methods in health research, but also to replace old structures with new ones. It's all about the big picture: digitization is now finally to become a matter of course in the healthcare sector as well.

Just under ten years ago, when physicians and researchers in Germany were pondering how to make the increasingly complex world of healthcare data accessible, they were looking enviously across the Atlantic and at our European neighbors. After all, they had been pushing digitization in healthcare for much longer, with more courage and commitment. *"Either we find a way, or we make one"*, became the motto of the founding fathers of medical informatics. At the time, it was unthinkable that university medicine would actually become the engine of digitization in the healthcare sector and set the pace, where government-mandated paths were followed rather sluggishly. With the opening of the German Portal for Medical Research Data, data from university medicine is now being made available to the public and industrial research community. Almost ten years of infrastructure development combined with the preparation of the regulatory basis for the use of health data for research purposes were in part agonizing preliminary work here, which is now finally bearing fruit.

Even if we can justifiably claim the digital development of health data from healthcare as a milestone, there is still much to be done. Preparations for a European Health Data Space are in full swing, and perhaps for the first time Germany is not just standing on the sidelines and watching. On the contrary, health data spaces are already being built in the context of Gaia-X, and with two German projects, we are playing a leading and formative role in Germany. Data from primary care is to be linked with data from nursing care, with self-measured data from smart devices and with self-collected data. Patients will thus increasingly become actors and designers in the health data space. Climate and environmental data will be interconnected with health data, enabling us to understand the influence of the individual environment on disease development and progression. We live in exciting times.

"While they were researching, x-raying, filming, radioing, the most delicious invention was born by itself: the detour as the shortest connection between two points", said Erich Kästner. So the digitization of health is finally arriving in Germany by a roundabout way – belatedly, to be sure, but not too late. On the way to the Health Data Space, we in Germany have perhaps found the shortest route.

I hope you enjoy reading the 15th issue of gesundhyte.de.

Yours

Roland Eils

Editor in Chief gesundhyte.de

Photo: Sebastian Runge

index

SCIENCE COMMUNICATION:

"very important for researchers is comprehensibility"

Interview with Beatrice Lugger, Director of the National Institute for Science Communication (NaWik) in Karlsruhe, Germany.

Science should be universally understandable. Good communication helps researchers to present their topics to the public. To do this, they need to translate their findings into everyday language and at the same time maintain the core message. However, many scientists lack the basic knowledge needed for a dialog with the population. Beatrice Lugger and her team at the National Institute for Science Communication (NaWik) in Karlsruhe have set themselves the task to improve the communication skills of scientists.

gesundhyte.de: *How do you define "science communication"?*

Beatrice Lugger: Opinions are still divided on what exactly is meant by science communication. First, there is internal science communication, i.e. communication within the scientific community. Secondly, there is so-called external science communication, which is directed toward the public and the media. And in external science communication, there are again two groups: on the one hand, there are the researchers and communicators who communicate from within science. They are part of the research system and cannot be objective.

"Opinions are still divided on what exactly is meant by science communication."

That's why we need a complementary view from the outside, from science journalism, which can classify, criticize and ask questions about the background. Often, other actors are included in external science communication, such as lawyers, various stakeholders, the scientific community, the economy, citizens, but also populists.

gesundhyte.de: *What tips do you give scientists at NaWik for science communication?*

Beatrice Lugger: On the one hand, we have identified five dimensions that are essential in preparation for any communicative situation. These are the goal, the target group, the medium, the style and the topic. At NaWik, we have put these in the form of an arrow, which illustrates that all five are interdependent and interrelated as well. For example, if the medium allows more scientific details, I can get a little more specific with the topic. But if my goal is to get the audience excited about science, too many facts and figures as style elements are potentially more of a hindrance. Therefore, it's important to be aware that every communicative situation is different, and you can never assume that you have everything under control just because you can communicate in principle.

Another very important factor for researchers is comprehensibility, comprehensibility, comprehensibility. A good advice is to come up with catchy key messages before an interview and test them in advance with friends and acquaintances, i.e.

Beatrice Lugger and her team at the National Institute for Science Communication (NaWik) in Karlsruhe have set themselves the task of promoting the communication skills of scientists (Photo: © NaWik).

with scientific laypeople. This helps to identify which core message is understood and can convey a topic particularly well.

gesundhyte.de: *What other aspects are particularly important in science communication?*

Beatrice Lugger: Scientists in particular should know how the media work. That can also mean knowing: "I don't have to give every interview". And of course, you can make mistakes: for example, overselling a topic, i.e. promising more in terms of potential solutions than the state of research actually says. Another perspective is central and part of the NaWik arrow: Communication must always have the target audience in mind. Another aspect that I would definitely like to mention is listening. Communication always means having an open ear, because that's the only way to find out what the other person is thinking and, at best, to be able to respond to that in your communication.

gesundhyte.de: *Can anyone learn science communication?*

Beatrice Lugger: Not everyone has to be in front of the camera later on, but I think it's essential to know how science communication works, who the actors are and how they act, and how the coordination with your own communication department works. That's why I recommend that all researchers get to grips with the basics of science communication. As in life, not everyone is talented in everything, but all scientists can and should familiarize themselves with the basics.

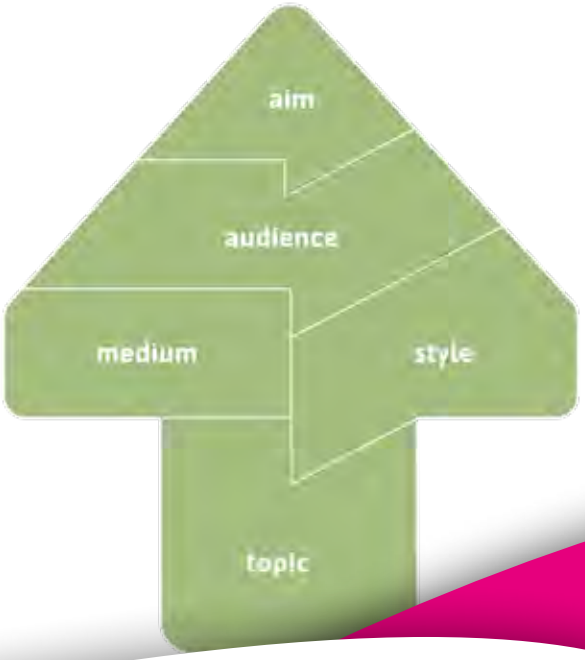


Figure 1: The NaWik arrow visualizes the five central dimensions of science communication.
(Source: © NaWik).

"Communication always means having an open ear, because that's the only way to find out what the other person is thinking and, at best, to be able to respond to that in your communication."

gesundhyte.de: *What opportunities does science communication offer scientists?*

Beatrice Lugger: On the one hand, you get an assessment of what other people think of your own research, especially in dialogue formats of communication. Another benefit is the recognition, the reputation, which can be associated with increasing visibility of one's own research. Thirdly, communication about science can create a somewhat better understanding in parts of society of how research actually works and that progress often has to be worked out slowly and laboriously in a process of steps forward and steps back.

gesundhyte.de: *What risks should scientists be aware of in their communication?*

Beatrice Lugger: In recent years, researchers have increasingly been publicly defamed and even threatened. Even if you as a scientist are not active in social media, you can be found - for example via your email address. You need a thick skin to deal with such insults and threats. On the one hand, there is the plan by "Wissenschaft im Dialog" (Science in Dialogue) with the Federal Association of University Communication to create a contact point for such situations. On the other hand, we at NaWik have introduced a Mayday button within the WissKon network, the network for communicating scientists: This button can be used to ask for help and advice from the WissKon community. Of course, you should also always inform your own press office in such situations and can ask for advice and support.

Leif Erik Sander
On his own communication experiences during the pandemic:

"I feel that science communication was very important during the pandemic and it still is. On the other hand, all communication during the pandemic took place under the circumstances of very intense social debates and disputes, and therefore, transfer of information was not always easy. Even when conveying neutral facts, statements were often construed by the media and the public, tainted with opinions and certain implications. If you compare this with science communication in pre-pandemic times or about non-pandemic topics, for example in cancer medicine, then these were very special circumstances for science communication. Nevertheless, even in situations with increased societal pressure on research and science, it is important to simply and clearly communicate the information that is available in a timely manner. When new information becomes available, it is also important to correct previous statement or to provide updates. Looking back, this was an intense and at times difficult task. But overall, I think it was important that even in the difficult situation of the pandemic, the scientific community tried to keep the public informed and to involve them in scientific debates."

Leif Erik Sander is Professor of Infectious Diseases and Director of the Department of Infectious Diseases and Critical Care at Charité and leads the Research Group for "Personalised Medicine in Infectious Diseases" at the Berlin Institute of Health at Charité (BIH).

Photo: © Wiebke Peitz | Charité

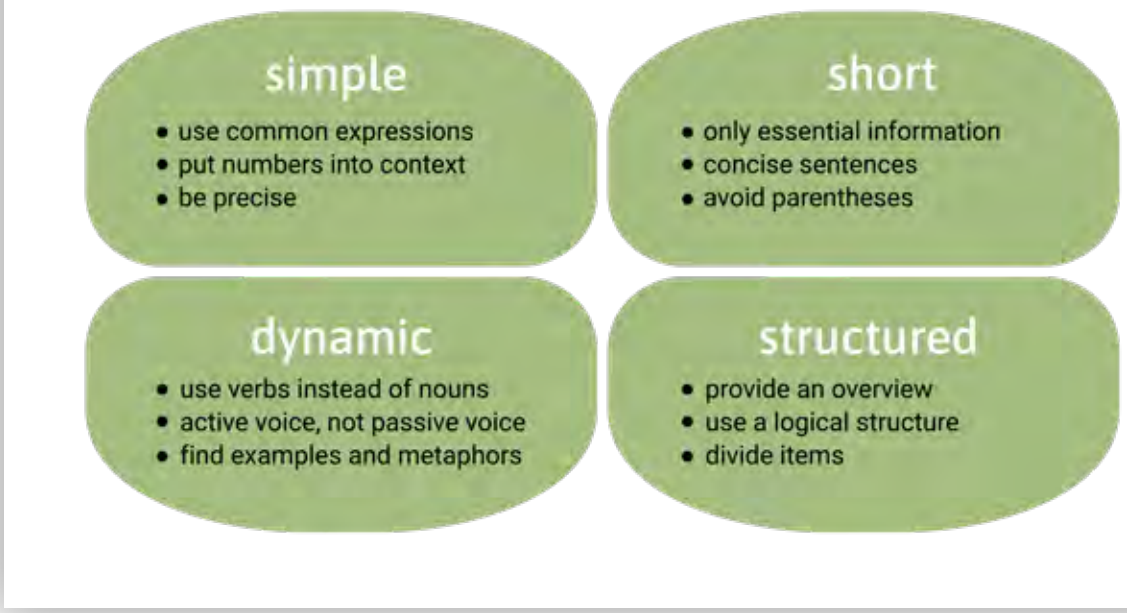


Figure 2: Simple, short, dynamic and structured, the NaWik cloverleaf summarises the rules for comprehensible communication (Source: © NaWik).

gesundhyte.de: *How do you become a good science communicator?*

Beatrice Lugger: It definitely makes sense to ask around in your own institute what training opportunities are available locally. Otherwise, of course, our courses at NaWik are worthwhile – from basic knowledge to social media, from presentation or interview and media training; the whole range up to science communication in funding applications. Plus a basic course is already available as an e-learning offer – also in an English version. It can be completed entirely online, independent of time and place, and is therefore ideal for integration into everyday research.

gesundhyte.de: *Which courses at NaWik are particularly popular?*

Beatrice Lugger: Basic courses and social media courses are at the top of the list at the moment. Overall, the demand for the different types of seminars is quite dynamic. In the last two years, the topic of visualization, for example, has increased strongly.

"Scientists in particular should know how the media work."

gesundhyte.de: *Why did you decide to pursue a career in science communication yourself?*

Beatrice Lugger: The main reason was that I am incredibly curious and did not want to commit myself thematically to a small field after my chemistry studies. Another is that I simply love communicating. You wouldn't believe it, I actually took a talent test at the employment office, and it also showed that I have a strong talent for communication. Then I just went my own way. For me, my current work also means constant own further training, and that's really cool.

The interview was conducted by Katharina Kalhoff.

Contact:
Beatrice Lugger
Managing Director National Institute
for Science Communication
lugger@nawik.de

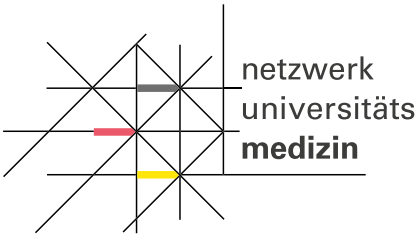
www.nawik.de

network of university medicine: institutionalization is the goal

Transition into a permanent platform funded by the federal government to develop and operate a multicenter research infrastructure

by Ralf Heyder

The Network of University Medicine (NUM) was founded in April 2020 as part of a project funded by the German Federal Ministry of Education and Research (BMBF), and as an ad hoc response to the first COVID-19 infection wave. Its mission initially was to coordinate national university medical COVID-19 research involving all disciplines and sites of university medicine (Figure 1). Patient care and pandemic management were to be directly supported with evidence generated or processed by the network as well as practical solutions. An important goal was to strengthen the pandemic preparedness of the health care and research system. This scope has recently been widened to research beyond COVID-19. To this end, NUM develops and operates research infrastructures (RIS) for the collection, storage and use of medical research data. An important partner in achieving this is the Medical Informatics Initiative (MII).



Permanent federal funding on the horizon

The NUM entered its first funding period on April 1, 2020, with traditional project funding. This quickly raised the question of how to sustain the cross-site RIS being created at NUM after the expiration of the initial project's funding period. Other federal project grants that focus on research infrastructures face similar challenges. One example is the MII.

Since January 1, 2022, NUM is continuing its work in a second funding period, which will end in mid-2025. Over the course of the approval process, the BMBF has expressed its intention to make the network permanent beyond the second funding period. This would be the first time that the federal government

Figure 1: The network partners in NUM (Source: Coordination Office Network of University Medicine).

Since the Network of University Medicine on COVID-19 links the experts of all 36 university hospitals, bundles their research activities and builds infrastructure for data sharing, future pandemics can be countered more effectively (Photo: AdobeStock © Robert Kneschke).

would provide indefinite funding for a structure that encompasses all of German university medicine. Therefore, the created structures have to be adapted for the intended continuation. **This implies three central fields of action:**

1. Opening up NUM's range of tasks and topics beyond COVID-19.
2. Reorganizing NUM governance to form more permanent networking, as it has so far been geared toward acute crisis management.
3. Further consolidating the RISs created in NUM, to enable "state of the art" research in different thematic areas.

Infrastructure for networked research as a core task

The RIS are currently undergoing consolidation and further development. This takes place in seven sub-projects. They form the so-called "infrastructure line" of NUM:

1. NUM Clinical Epidemiology and Studies Platform (NUKLEUS): This platform supports the standardized collection and provision of data, images, biospecimens, and information derived prospectively in the context of clinical (observational) studies. It is technically and conceptually based on the study platform of the German Center for Cardiovascular Research (DZHK).
2. NUM Data Integration Centers (NUM-DIZ): As of Jan. 1, 2023, funding for the DIZs (Data Integration Centers) transferred from the MII to NUM. This opens up the prospect for the DIZs to receive steady and thus permanent federal funding even beyond the end of the current third funding period of the MII, which ends in 2026 according to current plans. The further development of the DIZs will continue to be led by the proven structures of the MII.
3. NUM-Routine Data Platform (NUM-RDP): It aims at the (retrospective) extraction of treatment data on COVID-19 from the clinical information systems. The project is carried out in close cooperation with the MII and builds on its preliminary work by, among other things, complementing and linking the DIZs with a central data aggregation component.
4. Radiological Cooperative Network (RACOON): All university radiology departments in Germany have joined forces on this platform to collect radiological image and diagnostic data on COVID-19 in a structured manner across all sites. Due to this project, large data sets have been made available for joint research and the training of algorithms.
5. AKTIN Emergency Admission Registry (AKTIN@NUM): The registry existed prior to the establishment of NUM and was further expanded through NUM funding. It provides treatment data from the specific setting of emergency departments for research projects.

6. NUM Genomic Pathogen Surveillance and Translational Research (GenSurv): This platform collects sequencing and metadata of SARS-CoV-2 variants, e.g., to support surveillance for emerging viral variants using spread analysis.
7. National Autopsy Network (NATON): This service, expert and development platform collects data and biospecimens from autopsies and operates a registry for this purpose.

Modular research infrastructure development requires integration and harmonization processes

The heterogeneity of these sub-projects illustrates the main challenge that NUM and MII will have to solve together in the coming years. Medical research data are extremely heterogeneous. They go far beyond "real world data". Accordingly, the current NUM-RIS each addresses different types of data (for example, imaging or medication) and different ways of obtaining data (prospective vs. retrospective), as well as different data origination settings (emergency rooms, radiologies, etc.).

The approach of opening up this very broad and heterogeneous field via the development of RIS in several sub-projects working in parallel brings speed, but also involves risks. The sub-projects should not develop into data silos, which are not interconnected. Therefore, an umbrella structure is necessary that enables, among other things, record linkage, interoperability, and scalability via technical foundations that are harmonized as much as possible. A corresponding governance framework that supports the necessary integration and harmonization processes is currently being implemented in NUM.

A second governance challenge is to align the work of NUM and MII, which presently have parallel organizational and decision-making structures. **To this end, NUM and MII have reached the following key agreements as part of the last MII funding application:**

- Conduct joint training (e.g., IT security, consent management).
- Link the clinical community of NUM to the work in the MII, in particular
  - **(a)** for work on the core data set,
  - **(b)** for the implementation of data use projects.
- Sharing and further development of mutually supported infrastructures.
- Operation of joint MII and NUM working group(s).
- Cooperation and convergence at the level of governance of NUM and MII, in particular through mutual representation in relevant decision-making bodies.

An already visible implementation step in this cooperation is the establishment of a Coordination Group for Health Research Data Infrastructure (GFDI) jointly led by NUM and MII, which brings together NUM, MII and other relevant national initiatives from this field (e.g. German Centers for Health Research (DZG), National Research Data Infrastructure (NFDI), National Centers for Tumor Diseases (NCT)/National Decade against Cancer (NDK), but also actors from the field of government health telematics such as Gematik and BfArM). This body serves to coordinate overarching activities in order to exploit synergies and avoid duplication of effort.

Conclusion

In association with the MII and numerous other cooperation partners, NUM is excellently positioned to drive forward the development of RIS for networked biomedical research - nationwide and in the interest of all German university medical centers. However, this development needs endurance because Germany has to catch up considerably in terms of digitalization. The prospect of permanent funding from the federal government would provide the necessary continuity. It would also offer the opportunity to consolidate the infrastructure components established by the MII and NUM in an integrated governance and funding framework, supported by the entirety of German university medicine. To this end, however, the university medical centers must first create the prerequisites during the current NUM funding period. This includes, in particular, the close interlinking of NUM and MII.

Contact:

Ralf Heyder
Head of the Coordination Office
Network of University Medicine
Charité – Universitätsmedizin Berlin
ralf.heyder@charite.de

www.netzwerk-universitaetsmedizin.de

Photo: Wiebke Peitz I Charité – Universitätsmedizin Berlin

the german portal for medical research data FDPG

Access to Real World Routine Data of German University Hospitals

by Hans-Ulrich Prokosch, Julian Gruendner, Marie Gebhardt, Karoline Buckow, Philip Kleinert and Sebastian Semler

The German Portal for Medical Research Data (FDPG) was introduced in October 2022 as a central access portal to the data treasures of university hospitals.[1] Researchers from university medicine could access this portal. Since May 2023, access to the portal is also granted for researchers outside of German university hospitals.

With FDPG we achieved a major goal of the Medical Informatics Initiative (MII): a federated network across Germany that offers centralized access to data documented in the electronic medical records of all German university hospitals. This creates new cross-site data-based research opportunities. Currently, 31 German university hospitals have established local data integration centers (DIZ) and 26 are connected to this portal. Another eight DIZs will be established in the coming years. They, too, can enrich the data treasure trove in the future. Over the past five years, many hurdles have been overcome to establish the legal and organizational framework as well as the software components for a secure implementation of decentralized and centralized data use processes. In the following, we present the most important preliminary work that contributed to the success of the portal, introduce the current modules and functionalities of the FDPG, and address "lessons learned" from the first data use projects implemented.

[1] www.forschen-fuer-gesundheit.de

The challenges

The establishment of the FDPG has been accompanied by many standardization and harmonization activities in the Data Sharing, Consent, and Interoperability Working Groups established by the MII National Steering Committee. The first hurdle: Informed consent from patients in line with data protection requirements. Clinical data may only leave University Medicine sites in pseudonymized form if patients have agreed to a modular broad consent (MII Broad Consent, see also gesundhyte.de, issue 14, page 21ff). To legally secure the release of data and biospecimens, but also the decentralized provision of results of distributed evaluations, the working group "Data Sharing" coordinated regulations and structures for the use of patient data and biospecimens throughout Germany (uniform usage regulations, standardized data usage application and usage contract template). In the "Interoperability" working group, further hurdles were jointly removed. Here, technical specifications (so-called FHIR Implementation Guides) were developed for the basic and first extension modules of a Germany-wide coordinated core data set (person, case, diagnoses, procedures, laboratory, medication, consent, biospecimens and intensive care data).[2]

[2] www.medizininformatik-initiative.de/en/about-initiative/results

With the FDPG, the MII has achieved its most important goal: to support clinical and translational researchers in medicine and related sciences by providing access to clinical patient data from all German university hospitals via a central access portal that allows scientists to

- get an overview of the data and data types available in all DIZ,
- to precisely characterize their cohort for a planned research analysis,
- conduct feasibility queries to determine the extent of suitably available patient datasets,
- formulate data use applications to request data from all integrated DIZs and ultimately obtain the appropriate access to the demanded data for their research projects.

In the research projects, depending on the basis for use given under data protection law, either distributed evaluations can be carried out (the algorithm then "goes to the data", the data do not leave the sites) or central evaluations (if patient consent, i.e. Broad Consent, is available, the data can "go to the algorithm" and be made available for download).

Structure of the research data portal

The FDPG consists of three modules. First, a Germany-wide feasibility tool that builds on preliminary work from the CODEX project conducted jointly with the Network for University Medicine (Prokosch *et al.*, 2022a) and the MII project ABIDE_MI (Aligning Biobanks and Data Integration Centers Efficiently, Prokosch *et al.*, 2022b). This tool allows the selection of clinical patient parameters to characterize a cohort needed for a research project. Researchers can thus query across all university hospitals to determine how many patient records each site could contribute to the proposed project. It was further developed based on suggestions from MII working groups and successfully integrated into the overall FDPG framework.

The second module is a research project application management module. It was developed by a commercial software development partner (Appsfactory GmbH). The third module is a transparency register. This is a website designed to make the research projects conducted visible and understandable for patients. Individual FDPG software modules for the connection between the central research data portal and the data-supplying data integration centers are based on consortium concepts such as the Data Sharing Framework DSF (Hund *et al.*, 2021) or other

Figure 1: Feasibility portal, patient information and project registry are three relevant components of the German Portal For Medical Research Data. Information for patients and project registry are so far only available in german language.
(Source: https://forschen-fuer-gesundheit.de)

Figure 2: From research idea to publication: the career of a research project in the FDPG.
(Source: https://forschen-fuer-gesundheit.de).

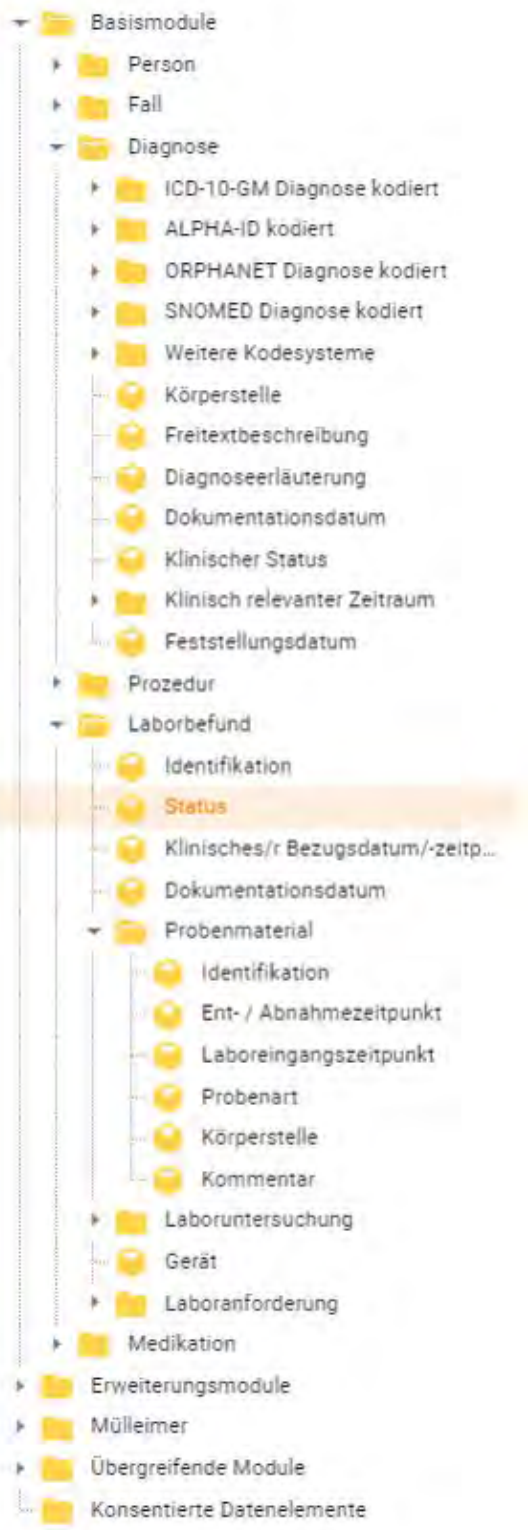
projects such as the AKTIN emergency admission register (Xu *et al.*, 2017) and have been integrated into the FDPG ecosystem for secure communication between the participating network partners (Gruendner *et al.*, 2022).

The local regulatory foundations and technical infrastructures have been established at the 31 German university hospitals participating to date in an iterative process within their DIZ. This includes the establishment of the associated data (and biospecimen) use and access committees (UACs) as well as the local trusted third parties.

The former examine the applications for use received according to organizational (feasibility), (data protection) legal and scientific aspects and decide on the participation of the institution in a user project. The latter protect the relationship of patient data and, if applicable, biospecimens to the patients concerned. To test the practicality and performance of the FDPG, an MII-wide projectathon was conducted, starting in September 2022:

Using four prepared research projects submitted to all DIZs through the FDPG, concepts, processes, and tools for data sharing were evaluated. The further steps of an overarching data use such as application review by the local UACs, data preparation, delivery, and provision - both for distributed evaluations and for central analyses - were also fully mapped in this projectathon. All participating 31 German university hospitals were asked to provide data for the projects.

30 DIZs participated, with an average of over one-third of DIZs (39 percent) in each of the four projects reporting back that they were providing data. The main reasons for not participating in the project were insufficient patient consent and incomplete implementation of the MII core data set modules at that time. The associated data use agreements were coordinated and handed to the participating sites for signature. The first project started in January 2023, analyzing the association between atrial fibrillation and a biomarker, NTproBNP value in clinical practice.



Subsequently, the other Projectathon projects started, which are currently sending out their data and analysis results. The Projectathon proves that all steps for the implementation of nationwide data use projects based on the routine data of university hospitals can be supported automatically via the FDPG, starting with feasibility inquiries, project/data application and the necessary approval processes as well as the final data provision, project implementation and project publication. However, it also shows that some process steps can still be improved. For example, FDPG employees had to finalize project applications via time-consuming communication with the applicants. This was mainly due to previously non-standardized specifications and missing FAQs and information in the FDPG web interface. For example, the correct use of free-text entry fields for project proposals was unclear.

This led to queries from the sites' DIZs and UACs, particularly regarding the selection of required data. Similar problems were encountered with the user interfaces for DIZs and UACs. The processes implemented were still new to users and needed to be explained to all stakeholders as they went through the process. The integration of decision-making processes into local DIZ work was successful, but needs to be streamlined to handle the expeditious processing of even large volumes of project applications. The paper-based collection of signatures, that is currently still required, is time-consuming given the large number of contractual partners and represents a considerable burden for all involved.

Outlook

In addition to these small improvements in the process flows and the specification of information for all process participants, technical enhancements will also contribute to the optimization of the FDPG from 2023. In a future version of the FDPG framework, for example, we will provide users with a hierarchical data catalog that enables precise and structured selection of the required data elements to avoid misunderstandings between requesters and data providers. Automated data extraction and preparation routines based on this will harmonize and accelerate the process of data provision at the sites.

Figure 3: Exemplary hierarchical catalog of data elements whose provision can be requested for a data use project.
(Source: Screenshot aus ArtDecor; https://art-decor.org/ad/#/mide-/datasets/dataset/)

In the MII expansion and extension phase, further development and integration as well as professional operation of all technical components of the FDPG ecosystem will now continue via new MII infrastructure projects (FDPG+, SU-TermServ, Transit, DSF-Community, 4C4MII). Further central consulting and service projects within the MII will contribute to the professional use of the established structures and tools for scientific knowledge gain, to the rapid integration of project results into patient care in the future, and to the further education of the medical informatics community as well as of physicians and researchers (for the use of all these tools). The cooperation project of the MII with the University Medicine Network aims to further expand the collaboration of both nationwide associations and to achieve further synergy effects by integrating relevant structures (See also previous article in this issue on page 12).

At the same time, a variety of new requirements will emerge from new clinical use case projects as well as from the use of the tools by researchers and from our increased involvement of patient representatives, which will be repeatedly addressed, evaluated, prioritized and implemented step by step by all stakeholders of these central infrastructure projects over the next four years from 2023 to 2026.

At the end of the MII set-up and networking phase, the provision of the FDPG impressively demonstrated how, despite all local heterogeneities, German university hospitals have agreed on common standards, tools and procedures over the past five years in order to support medical research in Germany with their data assets and to jointly contribute to the translation of findings into care. This seed has now sprouted. In the coming years, we will work together in the MII to ensure that it becomes a well-tilled field with a rich harvest.

References:

Prokosch, HU., Bahls, T., Bialke, M. *et al.* (2022a) The COVID-19 Data Exchange Platform of the German University Medicine. Stud Health Technol Inform; 294:674-678. https://ebooks.iospress.nl/doi/10.3233/SHTI220554

Prokosch, HU., Baber, R., Bollmann, P. *et al.* (2022b) Aligning Biobanks and Data Integration Centers Efficiently (ABIDE_MI). Stud Health Technol Inform; 292:37-42.

Hund, H., Wettstein, R., Heidt CM., Fegeler C. (2021) Executing Distributed Healthcare and Research Processes – The HiGHmed Data Sharing Framework. Stud Health Technol Inform; 278:126-133.

Xu, T., *et al.* (2017) Secure Electronical Communications and Data Transfers in a Clinical Environment. Stud Health Technol Inform; 245:1315.

Gruendner, J., Deppenwiese, N., Folz, M., *et al.* (2022) The Architecture of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) Patient Data Repositories: Design and Implementation Study. JMIR Med Inform;10(5):e36709. doi:10.2196/36709.

Contact:

Prof. Dr. Hans-Ulrich Prokosch
Chair of Medical Informatics,
Dept. of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität Erlangen-Nürnberg
ulli.prokosch@uk-erlangen.de

Dr. Julian Gründner
Chair of Medical Informatics,
Dept. of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität Erlangen-Nürnberg
julian.gruendner@fau.de

Marie Gebhardt
TMF – Technology, Methods, and Infrastructure for Networked Medical Research
Marie.Gebhardt@tmf-ev.de

https://forschen-fuer-gesundheit.de/

Photos: M. Rabenstein, Universitätsklinikum Erlangen; Simon Krikava; TMF e.V.

health data and research: how to combine the two?

Junior research groups in the medical informatics initiative

by Britta Sommersberg

The medical informatics initiative (MII) of the Federal Ministry of Education and Research (BMBF) creates the basis for an important "raw material" of modern medicine – clinical care and research data – available and usable for science. This applies, for example, to care data from cancer treatment or the results of diagnostic imaging procedures. All these data treasures can contribute to health research directly and noticeably improving diagnoses and therapy – provided the data are usable and available. Many clever minds are working and tinkering on this. And it is precisely these who are always in demand. Especially at the interface between medicine and informatics, the demand for young scientists is very high. For this reason, funding of junior researchers has always been a high priority in the BMBF's project funding in health research.

One new professorship = one junior research group

The BMBF supports all medical informatics consortia in recruiting scientific experts. For example, it is very attractive for university hospitals to appoint a new medical informatics professorship because they can apply for funding for a junior research group at the same time. Every consortium has made use of this option: 21 junior research groups have been launched since 2020. In total, the BMBF supports these junior research groups with 30 million euros. We are pleased to present four of these junior research groups below – one from each consortium. They show the entire thematic and methodological range of medical informatics, be it software platforms for molecular tumour boards, the integration of multimedia files such as images into existing electronic systems, new methods for earlier diagnosis of sepsis or the user-friendly development of digitally supported decision-making systems.

But take a look for yourself:

The four junior research groups at a glance

MoMoTuBo

The junior research group **MoMoTuBo** aims to develop concepts and prototypes for a software platform to support the processes of a molecular tumour board. What is new here is the modular character for the different sub-processes within the entire workflow and the focus on how reproducibility, documentation and update mechanisms are realised along the workflow. The junior research group is part of the MII consortium DIFUTURE and is funded by the Federal Ministry of Education and Research BMBF for five years with around 1.3 million euros.

Project lead: Dr. Zaynab Hammoud

IMPETUS

The **IMPETUS** junior research group aims to expand the Medical Data Integration Centre, which was set up in the HiGHmed consortium at the University Hospital Schleswig-Holstein in the MII, to enable the integration and use of as many multimedia objects and findings as possible, regardless of the format in which they are available. The junior research group is funded by the Federal Ministry of Education and Research BMBF for four years with about 1.5 million.

Project lead: Prof. Dr. Björn Schreiweis

CDS2USE

The junior research group "**CDS2USE**: Prospective user-friendly design of clinical decision support systems in the context of personalized medicine" aims to design technological innovations, which are developed for example in the medical informatics initiative, in such a way that they are user-friendly and context-adaptive. This is an elementary prerequisite for the developed innovations to be accepted and used in working practice in the future. The junior research group is part of the MII consortium MIRACUM and is funded by the Federal Ministry of Education and Research BMBF for five years with around 1.4 million euros.

Project lead: Dr. Brita Sedlmayr

MicrobiomSepsisPred

The junior research group "Prediction of sepsis based on microbiome sequence data (**MicrobiomSepsisPred**)" aims to develop methods to detect sepsis earlier. By means of sequencing and analysis of microbiome data, molecular signatures are sought that are suitable for the early detection of sepsis or the characterisation of sepsis risk. In addition, the junior research group aims to establish tools for the detection of antibiotic resistance profiles in the gut and skin microbiome of patients. This should create the basis for earlier and more targeted treatment of sepsis. The junior research group is funded by the Federal Ministry of Education and Research BMBF for four years with around 1.5 million euros.

Project lead: Dr. Ivana Kraiselburd

Contact:

Dr. Britta Sommersberg
DLR Project Management Agency |
Department Health Clinical Science and Digital Health
Bonn
britta.sommersberg@dlr.de

https://projekttraeger.dlr.de/en

Photo: private

MoMoTuBo: modular platform for molecular tumor boards

A new junior research group as part of the development and networking phase in the medical informatics funding concept

by Zaynab Hammoud

A tumor board is an interdisciplinary body that coordinates treatment plans for cancer patients. In the molecular tumor board (MTB), therapy decisions for cancer patients are made on the basis of various data. In the project MoMoTuBo (Modular Knowledge- and Data driven Molecular Tumor Board) the current approaches of MTBs, the necessary infrastructure, the processing and analysis of data, the linkage with prior knowledge and the integration into clinical practice as well as the current software support of MTBs are analyzed. On this basis, requirements and concepts for a modular design of MTB systems will be developed.

MoMoTuBo
Modular Knowledge- and Data-driven Molecular Tumor Board

In the MoMoTuBo project, the first step is to conduct a needs analysis for an MTB at the Augsburg University Hospital site. The further procedure includes an intensive literature search to

Figure 1: In a Molecular Tumor Board, the best therapy decisions for each single patient are made by a multidisciplinary panel (Photo: AdobeStock © Sharne T/peopleimages.com).

identify, analyze and evaluate existing MTBs at other sites and research groups. Further necessary requirements for MTBs will be identified and the processes and software of existing MTBs will be analyzed and compared. Since MTBs require working with knowledge and data that originate from different sources or from current and former patients respectively, internal sources are required. But also data from external sources like knowledge databases or scientific literature must be considered. Hence interfaces have to be developed to integrate and process this information. Based on the literature, we will determine, among other things, how these external and internal sources can be integrated into MTB processes or how an MTB can be meaningfully used in daily clinical practice. Based on this information, we will develop a concept for an MTB software system supported by various prototypical modules (Figure 1). Using machine learning algorithms, prototypical models can be created, especially for information integration into the software or for individual therapy discovery.

Contact:

Dr. rer. nat. Zaynab Hammoud
Junior Research Group Leader
of the Project MoMoTuBo
at the chair of "IT-Infrastructure for
Translational Medical Research",
Faculty of Applied Computer Science of the University
of Augsburg

www.uni-augsburg.de/de/fakultaet/fai/informatik/prof/misit/

Photo: Uni Augsburg

IMPETUS: Integration of Multimedia Objects

Junior research group investigates sustainable integration of multimedia objects in holistic knowledge management systems

by Björn Schreiweis, Hannes Ulrich and Björn Bergh

Imaging procedures are of crucial importance for patient treatment and research. However, only selected modalities (e.g. MRI, CT) are consistently stored in central picture archiving and communication systems (PACS) and evaluated by physicians in the radiological information system. The majority of the remaining modalities are either not stored at all in a sustainable IT system or are only available in departmental and special systems - then mostly not standardized, but in manufacturer-specific data formats. This also applies to other important clinical multimedia objects, such as curve displays of vital signs (e.g. ECG, EEG, RR) or audio files (e.g. from neurology), but also multimedia findings that contain both images and free text parts. Thus, a sustainable and reproducible reuse of these data for research studies is hardly possible.

The junior research group IMPETUS of the Medical Informatics Initiative addresses exactly these challenges. The goal is to

Figure 2: Today, many medical diagnostic images are only available in departmental systems. IMPETUS wants to change that (Source: DLR-PT/BMBF).

expand the Medical Data Integration Center (MeDIC) established in the HiGHmed consortium at the University Medical Center Schleswig-Holstein to enable the integration and reuse of multimedia objects and findings, regardless of format, storage or presentation, in a standardized form.

Pattern recognition methods improve the findability of image data

Pattern recognition methods (content-based image retrieval) can support medical staff and researchers in searching for relevant data. Semantic tagging using SNOMED CT, among others, is intended to annotate image data and thus make it easier to find. In this way, multimedia data can be searchable without loss of information and can be jointly evaluated. A major milestone has already been reached: the central PACS has been successfully and sustainably integrated with 6.6 million radiological image studies in MeDIC and thus made accessible for research.

Nevertheless, the semantic indexing of image data gives rise to our next research question: **How can established search procedures be applied equally to multimedia objects and vital signs?**

Contact:

Prof. Dr. Björn Schreiweis
Head of MeDIC UKSH
Head of Junior Research Group IMPETUS
Institute for Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein
bjoern.schreiweis@uksh.de
https://mi-ki.eu/en/welcome

Dr. Hannes Ulrich
Deputy Head of Junior Research Group IMPETUS
Institute for Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein
hannes.ulrich@uksh.de
https://mi-ki.eu/en/welcome

Prof. Dr. Björn Bergh
Chief Digital Officer University Hospital Schleswig-Holstein
Chairman Institute for Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein
bjoern.bergh@uksh.de
https://mi-ki.eu/en/welcome

Photos: private

CDS2USE: focus on user-centered design

Research for widespread and sustainable use of clinical decision support systems

by Brita Sedlmayr, Ian-C. Jung, Katharina Schuler and Maria Zerlik

Clinical Decision Support Systems (CDSS) can increase diagnostic accuracy and improve treatment decisions. However, medical staff use these systems too rarely in their daily clinical work. The reason for this is that CDSS are still developed with too much focus on technology and are largely geared to technical feasibility. The needs and requirements of medical staff and their specific working environment are not sufficiently taken into account. We, the Dresden junior research group CDS2USE, are developing solutions so that CDSS can actually improve care in everyday clinical practice.

Under the title "CDS2USE" (Prospective user-friendly design of clinical decision support systems in the context of personalized medicine), we develop concepts and tools for developers and researchers to involve potential users in the development process.

Our research covers the topics:

Context sensitivity

In order to map contextual information into CDSS algorithms, we identify factors influencing clinical decision making (e.g., time, case complexity) via systematic research. We model the clinical decision making process for selected use cases taking into account influencing contextual factors and validate these models with clinical staff.

Explainability of decisions

Decisions made by AI-based CDSS must above all be designed to be comprehensible. For this purpose, we collect concepts for interfaces that display explanations of CDSS results and derive design rules. Based on these rules, we develop and test new interface concepts for specific use cases. Such concepts can later serve as a blueprint for further use cases.

User-centered visualizations

In order to enable users to recognize correlations and trends in data in the best possible way, we develop visualization concepts for selected use cases based on existing work. We test these with users in order to provide a set of tools with user-centered graphical representations.

Figure 3: The team of the junior research group CDS2USE. From left to right: Brita Sedlmayr, Katharina Schuler, Ian-C. Jung, Maria Zerlik (Photo: Brita Sedlmayr).

Method Toolbox

We are developing an interactive toolbox with layman's instructions and best practice examples to provide developers and researchers with user-centered design and evaluation methods for CDSS. Our initial toolbox contains about 180 methods (e.g., the Pluralistic Walkthrough or the System Causability Scale), which are currently being converted into an interactive format.

CDS2USE opens up the possibility for us to explore user-centered design in a targeted way and with a lot of freedom in a dedicated team, thus making the topic more visible.

Contact:

Dr. Brita Sedlmayr
Head of the research area "Usability and Technology Acceptance" and head of the Junior Research Group "CDS2USE", Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

https://tu-dresden.de/med/mf/imb?set_language=en

Photo: MF/Stephan Wiegand

MicrobiomSepsisPred: "What triggers sepsis?"

New opportunity for early detection of sepsis by applying AI to microbial data

by Ann-Kathrin Brüggemann, Josefa Welling, Folker Meyer and Ivana Kraiselburd

The junior research group, "Prediction of sepsis based on microbiome sequence data" (MicrobiomSepsisPred), is funded by the German Federal Ministry of Education and Research. The group works on improving the diagnosis of sepsis in intensive care units. As patients in these units are at particularly high risk of sepsis, an early detection is crucial for a fast and effective treatment with appropriate antibiotics. For this, the analysis of a patient's microbiome, i.e. the totality of genetic material of the microorganisms in and on a patient, can play a significant role. The focus here is on changes in bacterial communities that differ significantly from natural fluctuations in the microbiome of healthy people.

In the past, lists of bacterial indicator species were used to classify health and disease. In contrast to this approach, we rely on complex models or time series analyses using machine learning. Here, we resort to methods of Deep Learning, a variation of machine learning that uses neural networks with a large number of inner layers. In Deep Learning, properties of the data (features) are "learned" from the data in "feature learning". The features are then used in artificial neural networks to analyze new data.

With this method, we were able to achieve good results, which already exceed the results of linear state-of-the-art models. A critical change in the bacterial community could signal an impending problematic development and represent a first indication of possible sepsis. This shows: Time series prediction on microbial data is a key building block in the development of models that can be used to predict and quantify sepsis risk.

We base our studies on metagenomic data obtained by sequencing specific regions of the prokaryotic 16S ribosomal RNA gene (16S rRNA). These gene regions contain important taxonomic information. Due to the low cost, rapid implementation (sequencing and analysis within 24-48 hours of sample collection), and widespread use of 16S rDNAg analysis, this approach has the potential to support clinical applications in the future.

Overall, our project offers an excellent opportunity to improve patient care by combining microbiology and AI.

Contact:

Dr. Ivana Kraiselburd
Junior Research Group Leader
"Prediction of sepsis based on microbiome sequence data" (MicrobiomSepsisPred)
Data Science group at the Institute
for Artificial Intelligence in Medicine (IKIM)
University Hospital Essen
ivana.kraiselburd@uk-essen.de

https://www.ikim.uk-essen.de/groups/ds

Photo: PicturePeople Fotostudio Essen

Figure 4: MicrobiomSepsisPred junior research group from left to right: M.Sc. Josefa Welling, PhD student; M.Sc. Ann-Kathrin Brüggemann, PhD student; Dr. Ivana Kraiselburd, Junior Research Group Leader (Photo: Data Science group, IKIM).

"POLAR research" with a difference

How polypharmacy and drug therapy safety are analyzed in the German Medical Informatics Initiative (MII)

by Daniel Neumann, Miriam Kesselmeier, Torsten Thalheim, Renke Maas, André Scherag and Markus Loeffler

Polypharmacy is the simultaneous and permanent use of several (usually five or more) medications. This is also referred to as multimedication or polymedication. The frequency of polypharmacy increases with age and with the number of diseases. The more medications taken at the same time, the more likely it is that one of them will cause an adverse effect or that it will interact with other medications. Therefore, polypharmacy increases the likelihood of drug risks. In this way, other diseases and additional therapy needs can be triggered. With better medication management, these effects would be partially avoidable.

The MII project POLAR_MI (https://www.medizininformatik-initiative.de/en/POLAR, "POLypharmacy, drug interActions and Risks") is a multidisciplinary consortium of stakeholders from the fields of medical informatics, biometrics, epidemiology, clinical pharmacy, clinical pharmacology, pharmacy and health research from 21 institutions and 13 university hospitals. POLAR_MI has demonstrated that data from the data integration centers (DIZs) of all four MII consortia can be used through standards and analysis scripts defined across consortia to jointly identify and analyze medication-related problems. The experience gained from POLAR_MI, which is based on "distributed analyses" using retrospective data from the DIZ, are used in the INTERPOLAR ("INTERventional POLypharmacy – drug interActions – Risks") project in the expansion and extension phase of the MII which has been started in January this year. In INTERPOLAR, algorithms will be developed that identify patients at particularly high risk for medication errors and side effects and alert ward pharmacists to potential risks. This can protect patients from potential medication risks. An intervention study will investigate whether this procedure leads to more efficient detection of medication errors and side effects of medications and their correction or avoidance compared to the current routine.

Polypharmacy as a growing problem in care delivery

Polypharmacy is a global health problem defined in 2019 by the World Health Organization (WHO) as the "concurrent and regular use of five or more over-the-counter, prescription, or traditional medicines" (see: Medication safety in polypharmacy: technical report[1]). The active ingredients contained in the medicines may influence each other. However, a clear, direct relationship between polypharmacy and a worsening of clinically relevant outcomes has not yet been demonstrated. One reason for this could be incomplete data collection in patients' electronic records.

Estimates of the prevalence of polypharmacy vary widely, ranging from 25 to 80 percent, depending on definition, geographic region, study design, and health care setting. In Germany, more than one-third (42 percent) of people over age 65 with public

[1] https://www.who.int/publications/i/item/WHO-UHC-SDS-2019.11

health insurance take more than five medications. One in four people have been prescribed potentially inadequate medications. There are many reasons for this: patients often simply forget to take prescribed medicines, which can have a negative impact on the course of the disease. Or the intake and documentation do not match (Schmiemann *et al.*, 2012).

Piece-by-piece indexing of inpatient treatment data with the help of DIZ

In the MII's POLAR_MI collaborative project, part of this outlined problem was addressed based on the (inpatient) treatment data available in the DIZs. Statistical analysis plans were created based on five pharmacological questions using only the FHIR resources of the basic modules of the MII core data set (https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set) defined by the MII's Interoperability Working Group. In doing so, the questions were designed to be more complex in their requirements for the data to be used. Despite predefined definitions in the core data set structure in the MII, it became clear in POLAR_MI that even at university hospitals a very large heterogeneity of the local organization of medication supply and its digital documentation had to be taken into account. However, the same was true for other data elements of the core data set structure; e.g., questions requiring day-specific time stamps to enable temporal classification of events (e.g., prescription of a drug and documented adverse drug reactions such as bleeding).

Plausibility check and original data comparison of inpatient treatment data

One of the five pharmacological questions was about the prevalence of the administration or prescription of amitriptyline during inpatient stays of patients who were 65 years or older. This is a tricyclic antidepressant prescribed for the treatment of depression and chronic pain. The medication is declared as *"Potentially Inadequate Medication"* (PIM) for persons over 65. The period 2018-2021 was defined for the retrospective review. The meta-analysis on preliminary local results from eleven sites showed that this PIM was identified in 1,390 cases from 136,780 inpatient stays. Even taking into account the heterogeneity between the sites, this corresponds to a proportion of approximately one percent. However, the heterogeneity between the sites was very large. This heterogeneity can be partly explained by the fact that not all clinics at all sites were connected to the system and that some sites could not yet cover the entire evaluation period. Furthermore, there was a large proportion of cases for which no information on inpatient medication was available. In total, 406,682 inpatient stays of a total of 251,826

Figure 1: Virtual INTERPOLAR kickoff (Source: SMITH office with INTERPOLAR project coordination).

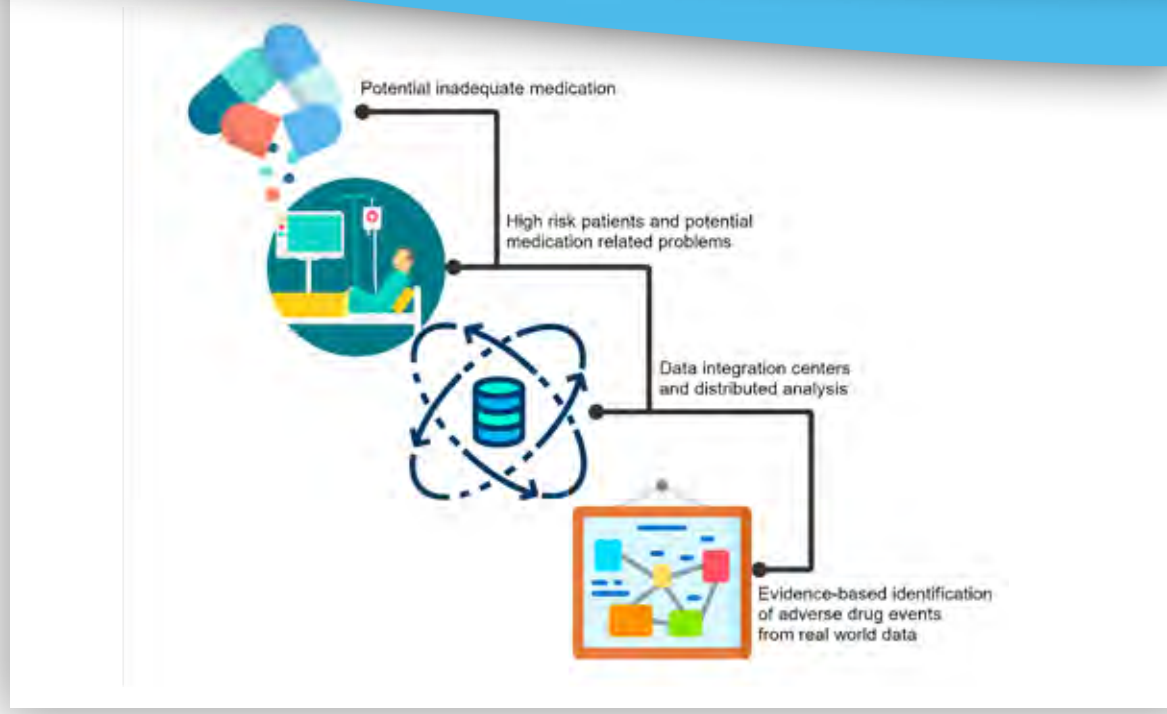


Figure 2: IT-supported care can prevent inadequate medication in high-risk patients
(Source: Own representation, this representation is designed with images from Flaticon.com).

patients fulfilled the formal inclusion criteria. The discrepancy between those and the assumed available cases can be explained, at least in part, by the affiliated clinics of the sites, because not all are equally digitised. In order to examine the plausibility of the data derivation, an extensive original data comparison was carried out locally at all sites across all subprojects, while maintaining data protection. A total of 1,300 files of patients with inpatient stays in which an event under investigation (e.g. certain PIM) was identified were reviewed, as well as control cases without an identified event. We found that, depending on the question, the proportion of stays falsely classified as positive with an event was between 5 and 15%. And conversely, we can estimate that the overlooked events are of the same order of magnitude.

Drug therapy safety with data-based risk scoring

In order to address polypharmacy in the sense of improving drug therapy safety, lists of *"Potentially Inadequate Medication"* (PIM) such as the PRISCUS list (see also: Priscus 2.0[2]) or the FORTA list (Fit fOR The Aged) were integrated into POLAR_MI, but also information on contraindications from specialist information was made accessible for digital analyses. However, it quickly became clear that PIM and contraindication information is often not available in a time-related manner with concomitant time stamps or is not sufficient on its own.

This is precisely where the new INTERPOLAR project comes in. The goal is to identify high-risk patients. This is made possible by a daily automated screening of routine care data of inpatients, which outputs a risk score for each individual. This is displayed to ward pharmacists when reviewing medication plans. All risk assessment information used for the risk score result should be made available to the ward pharmacists (*"human in the loop"* approach to determine sensitivity while avoiding over-alerting). In this way, the ward pharmacists retain the decision-making authority as to which patients they prioritize on the topic of drug therapy safety.

INTERPOLAR studies

To demonstrate the clinical utility and transferability of the approach of IT-assisted care by ward pharmacists, we plan to conduct two linked studies to optimize care. First, an evidence-generating, confirmatory intervention study with a cluster-randomized approach will be conducted. It will compare the usual care provided by ward pharmacists (Usual Care) with IT-supported care in six wards in each of eight university hospitals. This study will be supplemented by a subsequent translational study with seven further university hospitals, which will only use IT-supported care and is intended to consolidate the results of the confirmatory study with further data from prac-

[2] https://www.priscus2-0.de/

tice. The two studies together are expected to include approximately 120,000 hospitalized patients.

INTERPOLAR extensions

The findings from POLAR_MI collected in the course of care are also to be expanded to include data with the participation of patients, the PROMs (*"patient-reported outcome measures"*). In addition to the collection of PROMs, INTERPOLAR also plans to include data from previous hospital visits based on *"broad consent"* (see gesundhyte.de Issue 14, page 21). By better temporal classification of events, a *"patient journey"* (in the sense of recording the course of care of patients) as complete as possible should be developed, on the basis of which further predictive and dynamic influencing factors can be researched.

References:

Schmiemann G, Bahr M, Gurjanov A, Hummers-Pradier E: Differences between patient medication records held by general practitioners and the drugs actually consumed by the patients. Int J Clin Pharmacol Ther 2012; 50: 614–7

Moflhammer D, Haumann H, Mörike K, Joos S: Polypharmacy—an Upward Trend with Unpredictable Effects, Dtsch Arztebl Int 2016; Vol. 113, No. 38, 627-633.

Güster C, Klose J, Schmacke N (eds.): Versorgungs-Report 2012. Schwerpunkt: Gesundheit im Alter. Stuttgart: Schattauer 2012; 111–30.

Fried TR, O'Leary J, Towle V, Goldstein MK, Trentalange M, Martin DK: Health outcomes associated with polypharmacy in community-dwelling older adults: a systematic review. J Am Geriatr Soc 2014; 62: 2261–72

Scherag, A, Andrikyan, W, Dreischulte, T *et al.* : POLAR – „POLypharmazie, Arzneimittelwechselwirkungen und Risiken" – wie können Daten aus der stationären Krankenversorgung zur Beurteilung beitragen?. Präv Gesundheitsf 2022; https://doi.org/10.1007/s11553-022-00976-8

Holt S, Schmiedl S, Thurmann PA: Potentially inappropriate medications in the elderly: the PRISCUS list. Dtsch Arztebl Int 2010; 107: 543–51

Mann NK, Mathes T, Sönnichsen A, Pieper D, Klager E, Moussa M, Thürmann PA: Potentially Inadequate Medications in the Elderly: PRISCUS 2.0—First Update of the PRISCUS List. Dtsch Arztebl Int

PROFILE: Research project POLAR_MI

POLAR_MI

The use case "POLypharmacy, Drug Interactions and Risks", which encompasses all four consortia of the Medical Informatics Initiative (MII), used MII methods and processes to contribute to the detection of health risks in patients with polymedication. The project was funded by the German Federal Ministry of Education and Research for three years (2020-2022) with a total of around 5.5 million euros.
https://www.medizininformatik-initiative.de/en/POLAR

PARTICIPATING PARTNERS:

M. Löffler (University Leipzig, PI),
A. Scherag (Co-PI) **& D. Ammon & M. Hartmann & K. Farker** (Jena University Hospital),
R. Maas & M. Fromm (Friedrich-Alexander-Universität Erlangen-Nürnberg),
U. Prokosch (Erlangen University Hospital),
U. Jaehde (Rheinische Friedrich-Wilhelms-Universität Bonn),
S. Zenker & M. Coenen (University Hospital Bonn),
W. E. Haefeli & H. Seidling (University Hospital Heidelberg),
J. Gewehr (University Medical Center Hamburg–Eppendorf),
T. Wendt & Y. Remane (University Hospital Leipzig),
T. Dreischulte & F. Albashiti & M. Andraschko (University Hospital LMU Munich),
A. Eisert & I. Lutz (University Hospital Aachen),
M. Boeker (University Hospital Freiburg),
A. Michel-Backofen & K. Marquardt (University Hospital Gießen),
D. Tiller (University Hospital Halle),
I. Cascorbi & B. Schreiweis & B. Bergh & C. Bulin (University Hospital Schleswig-Holstein),
M. Schwab (University Hospital Tübingen),
T. Ganslandt (University Heidelberg, Med. Fakultät Mannheim),
P. Thürmann (University Witten/Herdecke)

2023; Jan 9;(Forthcoming):arztebl.m2022.0377. doi: 10.3238/arztebl.m2022.0377. Epub ahead of print. PMID: 36507719.
Kuhn-Thiel AM, Weiss C, Wehling M, members Faep: Consensus validation of the FORTA (Fit fOR The Aged) List: a clinical tool for increasing the appropriateness of pharmacotherapy in the elderly. Drugs Aging 2014; 31: 131–40

Prof. Dr. Markus Löffler
Principal Investigator POLAR_MI and INTERPOLAR
Director of the Institute for Medical Informatics, Statistics and Epidemiology (IMISE) at Leipzig University

Contact:

Dr. Daniel Neumann
Institute for Medical Informatics, Statistics and Epidemiology (IMISE)
Leipzig University
daniel.neumann@uni-leipzig.de

Prof. Dr. André Scherag
Principal Investigator POLAR_MI and INTERPOLAR
Director of the Institute of Medical Statistics, Computer and Data Sciences at the Jena University Hospital

Photos: Neumann: private, Löffler: UKE, Ronald Frommann, Scherag: IMSID, UK Jena

PROFILE: Research project INTERPOLAR

INTERPOLAR

The cross-consortium project "INTERventional POLypharmacy – Drug Interactions, Risks" builds on POLAR. In a study at eight university hospitals, the usual care provided by ward pharmacists is being compared with the IT-supported care developed in POLAR. This will be supplemented by findings from pharmacological research, such as the PRISCUS list. If successful, this study will be supplemented by a translational study with seven other university hospitals, which will introduce only the IT-supported variant and lead to consolidation of the results of the confirmatory study with further data from practice. INTERPOLAR is funded by the German Federal Ministry of Education and Research for four years with approximately nine million euros.
https://www.smith.care/en/interpolar

PARTICIPATING PARTNERS:

M. Löffler (University Leipzig, PI),
D. Neumann (University Leipzig, Project Coordinator),
A. Scherag (Co-PI) **& D. Ammon & M. Hartmann & K. Farker** (Jena University Hospital),
R. Maas (Co-PI) **& M. Fromm** (Friedrich-Alexander Universität Erlangen-Nürnberg),
T. Ganslandt & F. Dörje (Erlangen University Hospital),
U. Jaehde (Rheinische Friedrich-Wilhelms-Universität Bonn),
S. Zenker & I. Schulze (University Hospital Bonn),
H. Seidling & T. Hoppe-Tichy (University Hospital Heidelberg),
M. Sedlmayr & H. Knoth (University Hospital Dresden),
S. Härterich & M. Baehr (University Medical Center Hamburg–Eppendorf),
T. Wendt & Y. Remane (University Hospital Leipzig),
T. Dreischulte & F. Albashiti & M. Andraschko (University Hospital LMU Munich),
A. Eisert & I. Lutz (University Hospital Aachen),
C. Haverkamp & M. Hug (University Hospital Freiburg),
A. Michel-Backofen & G. A. Jost-Buhl (University Hospital Gießen),
D. Tiller & R. Harnisch (University Hospital Halle),
K. Strauch & I. Krämer (University Hospital Mainz),
I. Cascorbi & B. Bergh & B. Schreiweis (University Hospital Schleswig-Holstein),
F. Nensa & T. Brieden & J. Schnurrer (University Medical Center Essen),
P. Thürmann & S. Schmiedl (University Witten/Herdecke)

digital hub LeMeDaRT – a space for health data

Lean medical data in the right place at the right time

by Joachim E. Fischer and André Baumgart

Highly complex industries such as aviation demonstrate that digitization across different areas of expertise and along value chains can not only optimize quality, performance, and safety, but also make certain offerings possible in the first place, such as the real-time availability of seats on a scheduled flight from Frankfurt to New York or the avoidance of waiting loops at the destination airport through predictive algorithms that even take the weather and wind into account.

In the healthcare sector, there are already today sector-specific, isolated individual solutions with a high degree of digital technology consolidation and singular use of data, such as in oncology. In everyday care outside of top university medicine, however, a mix of analog, manual, semi-digital and digitized practices persists with information disruptions, inefficiencies and suboptimal allocation of human resources.

Health systems of the future will use digital solutions to introduce patient-centered care pathways that distribute work tasks along the continuum from health promotion and prevention to cutting-edge medicine among different actors in multiprofessional teams.

In such a future connected network, how would the "patient journey" accompanying the complexity from the first treatment request to the final care look like?

Although many things in medicine are similar to standardized processes such as those in aviation, they differ in one essential respect: In many frequent treatment cases, there remains a residual uncertainty about the state of health and there is no conclusively correct therapy. In the Digital Progress Hub LeMeDaRT there is also the fact that in the selected use cases patients interact in complex ways with various service providers in the system with an open outcome of the treatment.

Decisions along the "patient journey" are always based on data as well as information newly created by events and context. How does this information need to be prepared so that it can provide clinical benefit outside of narrow expertise? How will it be possible in the future to provide all stakeholders with the necessary data and clinically useful information? How can those persons involved in treatment, i.e. patients and specialists, have access to exactly the right data at the right time in a format that facilitates treatment rather than making it more difficult? This is even more challenging in complex decision-making processes with uncertainty than in coding a billing case.

Figure 1: Digitizing existing processes alone is not enough. How can we redesign processes to incorporate digitization in a way that creates added value for everyone? (Photo: AdobeStock © xyz+).

It is therefore not surprising that, until now, billing and administrative processes have been digitized, but not core medical processes such as personalized anamnesis and treatment management that takes patients' preferences into account, including motivation and support along sometimes complex decision-making pathways. The Digital Progress Hub LeMeDaRT addresses this challenge in three exemplary use cases in collaboration between medical informatics, clinics and practitioners in three rural regions. These use cases involve relevant or frequent treatment events that require the collaboration of many stakeholders across sector boundaries. All use cases have in common that they can be solved analogously and without adequate information processing only with high, impracticable coordination efforts.

The three treatment occasions are:

1. preparation before a complex tumor operation,
2. preventive clarification of very frequent asymptomatic liver afflictions, and
3. a more precise differential diagnosis of transmissible respiratory diseases.

LeMeDaRT focuses on the benefits for the patients in the use cases.

Use case surgery

Every patient with an upcoming complex tumor surgery would benefit from having his or her health status optimized prior to surgery. This is because the state of health before surgery has a direct impact on recovery afterwards. Personalized and targeted mental and physical fitness training before surgery requires the interaction of specialists from surgery, anesthesia, physiotherapy, nutritional counseling, clinical psychology, and nursing. The use case solves how a digitally supported treatment pathway is designed and put into practice with the involvement of the affected person from indication to aftercare across all disciplines.

Use case prevention

Today's lifestyle leads to a gradual fatty degeneration of the liver in many people, which can significantly shorten life expectancy if it progresses. Depending on the definition of fatty liver, up to a quarter of Germans are affected. Some are

noticed during the health check-up at the family doctor with slightly increased liver values, others not at all. Today, there are technological innovations that allow reasonable screening. Digitization makes it possible to harness these innovations across sector boundaries in preventive treatment pathways. In this way, any late effects of fatty liver disease can be reduced.

Use case respiratory diseases

The nose begins to run, headache or sore throat and some fever. Behind it could be a dangerous variant of the SARS-CoV-2 virus, an incipient bacterial infection, influenza or a comparatively harmless cold. The pandemic has shown us how useful it can be to know what has just infected you. Digitization makes it possible to combine clinical information from current cases and random molecular biology tests in a timely manner. This would allow each affected person to match his/her own disease signs and locations against an up-to-date regional atlas and self-learning regionalized symptom and progression information. The result for affected persons as well as for physicians is a better assessment of the probability of the respective disease and behavioral recommendations derived from it. Since every further development of the course of the disease can be included in the assessment, the system becomes increasingly accurate. This is of practical importance in the case of diseases such as influenza, where, according to current knowledge, possible drug treatment should be started as early as possible when it is actually not yet really clear whether influenza is present.

The key: processes as interoperable chains of events

The challenge for LeMeDaRT in handling these use cases is first to define the event chains and store the information associated with each event. Who needs which data and when in the process? How should data be presented so that the information helps everyone involved and facilitates the next steps in the treatment or care path? It quickly becomes clear that building digital infrastructures of compatible interfaces between systems requires prior exact specification of event data streams. These must be broken down into smaller, self-contained and precisely defined individual elements, which contain not only a definition of the data, but also of the parties involved and the events associated with them.

The data itself must be described in an interoperable form, such as with ontologies like SNOMED or LOINC (see Issue 14, p. 25). Here, the Medical Informatics Initiative offers excellent starting points through its work on a core data set, as do the activities of the "Zentralinstitut der Kassenärztlichen Vereinigung" in the definition of Minimum Information Objects (MIOs). However, our work shows that annotation and semantic interoperability alone are necessary but not sufficient conditions. In addition, there is the dimension of visualization, temporal availability of information, and integration into concretely defined event streams along the treatment path. Conceptually, the LeMeDaRT project is a complex intervention in an existing system and context. Therefore, continuous implementation with feedback loops from the context and system is essential for later exploitation.

Translational research shows that a purely theoretically guided development in the university has no chance of succeeding in practice later on. Therefore, in our project the comprehensibility for patients as well as the contribution to the motivation to change an information event unit must also be taken into account, if data along the treatment pathway are also to provide clinical benefits. For the Medical Informatics Initiative, the project can generate added value by operationalizing interoperability in the periphery and between non-medical professions, and by providing exemplary intelligent solutions with concrete patient benefits for common everyday medical challenges.

Figure 2: LeMeDaRT subproject leader meeting (Photo: © Corinna Müller).

Contact:

Prof. Dr. med. Joachim E. Fischer
Professorship Medical Faculty Mannheim,
Leader of LeMeDaRT,
Heidelberg University,
Center for Preventive Medicine and Digital Health (CPD), Division of General Medicine, Mannheim
joachim.fischer@medma.uni-heidelberg.de

https://www.umm.uni-heidelberg.de/medical-faculty-mannheim/general-medicine/

Dr. André Baumgart
Head of AI-based Health Systems,
Co-Leader of LeMeDaRT,
Center for Preventive Medicine and Digital Health (CPD), Division Biomedical Informatics, Mannheim
andre.baumgart@medma.uni-heidelberg.de

https://www.umm.uni-heidelberg.de/medical-faculty-mannheim/general-medicine/

Photos: Fischer: Tobias Schwerdt; Baumgart: private

The research project in brief

LeMeDaRT is one of the six "Digital Hubs: Advances in Research and Health Care" funded since fall 2021. The Federal Ministry of Education and Research (BMBF) is providing a total of around 50 million euros until 2025 for this flagship initiative of its digital strategy. The task of the Digital Hubs is to allow the pioneering work of the Medical Informatics Initiative on digitization in medicine to flow from university hospitals – initially in pilot projects – into all areas of the healthcare system: from outpatient care in the general practitioner's office and inpatient stays in the local hospital to care in rehabilitation and nursing facilities.
https://www.umm.uni-heidelberg.de/miism/data-analysis-and-modeling-in-medicine/research-areas/lemedart

Federal Ministry
of Education
and Research

News from the BMBF

Digitization of health research enters new phase

Advancing digitization in medicine and improving medical care for people – this is what the Medical Informatics Initiative (MII), funded by the German Federal Ministry of Education and Research (BMBF), stands for. Starting in 2023, the MII entered its expansion and extension phase, which the BMBF will fund with around 200 million euros for four years. One of the central goals: Digital innovations from university medicine should flow as broadly as possible into the healthcare system and noticeably improve care. To this end, the MII partners are networking even more closely with each other, as well as with other important players and initiatives in health research and with regional partners – from local hospitals to rehabilitation facilities.

Noticeably improving care practice

Many of the MII's clinical use cases to date, which concretely demonstrate the benefits of digitization in medicine, will be extended in the consolidation and further development phase. Diagnostics and therapy of cardiovascular diseases, personalized cancer medicine, and chronic lung diseases (asthma, COPD) will continue to be the focus, as well as better protecting people from drug interactions (see article p. 27) and hospital-acquired infections. In addition, the MII is expanding its range of clinical use cases, for example, to include aspects of sleep medicine and ophthalmology that affect many people. In the coming years, methods will also be developed to scientifically assess the quality of life of patients – for example, people with anorexia, a transplanted kidney or permanent physical ailments. With this new use case, the MII aims to help ensure that modern health research and care focus primarily on people and their well-being.

Establishing a decentralized research data infrastructure

In its consolidation and further development phase, the MII in Germany is to become the driving force and impetus behind the development of a decentralized research data infrastructure for health data as anchored in the coalition agreement. Important building blocks of this infrastructure are the data integration centers (DIZ) of university medicine. Established by the MII, the DIZs were integrated into the national Network of University Medicine (NUM) at the beginning of this year. Routine and research data are thus brought together in a decentralized manner and made available to research centrally. This strengthens university medicine as a whole, and synergies can be exploited to the fullest (see article p. 12). Researchers who wish to use data and biospecimens from the MII for scientific questions have already been using the MII's German Portal for Medical Research Data (FDPG) as a point of contact since the beginning of this year (see article p. 15). However, the FDPG is not only intended to help researchers find the right data for their projects. The portal is also intended to create transparency among the general public. To this end, it informs interested citizens about all ongoing projects that conduct research with patient data from the MII.

Data protection and data security

Comprehensive data protection and data security are a central factor in the success of the MII. The volun-

Computer-aided analysis of complex data helps physicians to make precise diagnoses and tailor therapy decisions.

Photo credit: DLR-Projektträger/BMBF

tary and informed consent of patients is a prerequisite for using their data for research projects. The MII involves data protection officers, ethics committees and representatives of patient organizations in its planning. Proven IT experts ensure that patient consent is securely documented electronically and carefully managed. Patients can withdraw or change their consent at any time.

Medical informatics initiative – the key data

Consortia (2018-2026): Four MII consortia are creating the conditions for the cross-site use of data from research and patient care. The MII demonstrates the medical added value of these infrastructures for data analysis and patient care in concrete use cases. The BMBF is funding the consortia with a total of more than 400 million euros during the development and networking phase (2018 to 2022) and the currently ongoing consolidation and further development phase (2023 to 2026).

Digital Hubs: Advances in Research and Health Care (2021-2025): The six funded hubs incorporate data from regional care into MII structures and solutions in pilot projects. They demonstrate the benefits of this networking for regional patient care in practical use cases (see article p. 32). The BMBF is funding the Digital Hubs with 50 million euros.

Junior research groups (2020-2026): Well-trained specialists are a key factor in the success of the MII, which is why the BMBF is funding a total of 21 junior research groups with around 30 million euros (see article p. 20).

Further information is available at:

medizininformatik-initiative.de/en

Artificial intelligence to detect legionella tricks

In view of rising energy prices, it seems obvious to lower the temperature of hot water. However, anyone who wants to save costs in this way could be bringing a serious health risk into their home. Dangerous pathogens such as Legionella, which multiply particularly well in warm water, can lurk in drinking water.

Drinking and washing hands are usually not a problem. However, water droplets and spray mist, such as those produced when showering, can carry Legionella bacteria into the air and from there into the lungs. The rod-shaped bacteria are a particular danger for immunocompromised and elderly people: they can trigger severe pneumonia, known as Legionnaires' disease. Every year, up to 30,000 people contract the disease in Germany alone. In up to ten percent of cases, the disease is fatal. "There is still a lot of need for optimization in the treatment of Legionnaires' disease," says Professor Dr. Dominik Heider of Philipps University Marburg. The bioinformatician and his team are developing a software tool that should significantly improve the diagnosis and treatment of severe Legionella infections. The German Federal Ministry of Education and Research (BMBF) is supporting the project as part of the "Computational Life Sciences – CompLS" funding initiative.

More than 300 virulence factors

The research project focuses on the so-called virulence factors of the bacteria. They determine the pathogenic effect of the bacteria and decide how well the pathogen succeeds in penetrating the host body and surviving there. Legionella are particularly effective and tricky at this. As intracellular bacteria, they attack the very cells of the human defense system that are supposed to render them harmless. In this way, they are protected from their attacks and can at the same time multiply undisturbed inside the cells. In the course of co-evolution with their hosts, Legionella have developed more than 300 such virulence factors, often by adopting gene segments of the host into their own genome. This enables them to manipulate vital host cell functions and use them for their own benefit.

Bioinformatician Heider and his team are building a database with all known virulence factors of the total of more than 60 species of Legionella. "The background to this is that the different Legionella strains each have different strategies," explains Heider. With the help of the data collection, the researchers are training an artificial intelligence (AI)-based diagnostic tool. "In the future, this should be able to determine exactly which strain of Legionella a particular patient is infected with and which virulence factors with which function are currently active," Heider said.

AI analysis in real time

This information is crucial for the appropriate, individually tailored therapy for Legionnaires' disease. Because Legionella hides inside host cells, many common antibiotics are ineffective. There-

fore, it is important that the right drugs are used. In particularly severe cases, a mix of several agents is also necessary. In practice, the software tool could also save valuable time in finding the right medication. “Patient samples would no longer have to be cultured in the lab first, but would be analyzed by the AI in real time,” says Heider. “On this basis, physicians could then directly start with the most suitable treatment.”

The scientists expect their development to be ready for use in a few years. Nevertheless, they are thinking one-step further: if the software proves successful in clinical practice in the treatment of Legionella infections, it could also be used in the fight against other pathogens. After all, they all use virulence factors to successfully infect the host body. Thus, the tool could also reveal new targets for treating antibiotic-resistant bacteria. “The disease-causing mechanisms of the individual pathogens are indeed very different,” says Heider. “But our newly developed AI can, in principle, learn to recognize them.”

Further information is available at:

gesundheitsforschung-bmbf.de/de/digitalisierung-und-kunstliche-intelligenz-9461.php

They multiply well in warm water: the rod-shaped Legionella bacteria can cause dangerous pneumonia.

Photo credit: Adobe Stock / peterschreiber.media

Tracking down the causes of leukemia

The stem cells in the bone marrow are an important reservoir of the human body. They give rise to the numerous different types of cells that circulate in the blood. A stem cell always divides into two daughter cells, one of which normally develops into a mature blood cell, while the other remains in the bone marrow as a stem cell. In this way, several billion new blood cells are created every day, which take on important tasks such as immune defense (e.g. the white blood cells) or oxygen supply (the red blood cells).

Since stem cells serve as blueprints for future generations of blood cells, they must be particularly well protected. This is because errors in the genetic material would be passed on to the daughter cells and thus multiplied. Such defective blood stem cells are the starting point for blood cancer. Therefore, blood stem cells are usually in a kind of dormant state in a niche in the bone marrow. This special environment protects them from external influences and the dormant state is only lifted when they are supposed to divide there.

T cells can sort out defective stem cells

Nevertheless, errors in the genetic material occur time and again, especially during cell division. These mutated cells are recognized and sorted out via another protective mechanism, as the scientists of the LeukoSyStem consortium were able to demonstrate. “In the stem cell niche there are immune cells, so-called T cells, which regularly check whether the blood stem cells have dangerous changes. If so, they sort them out,” explains Dr. Simon Haas from the Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin and the Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC). He coordinates the consortium, which is funded by the German Federal Ministry of Education and Research (BMBF).

The T cells recognize the altered stem cells by means of certain signal molecules on their cell surface. This recognition mechanism has so far been described

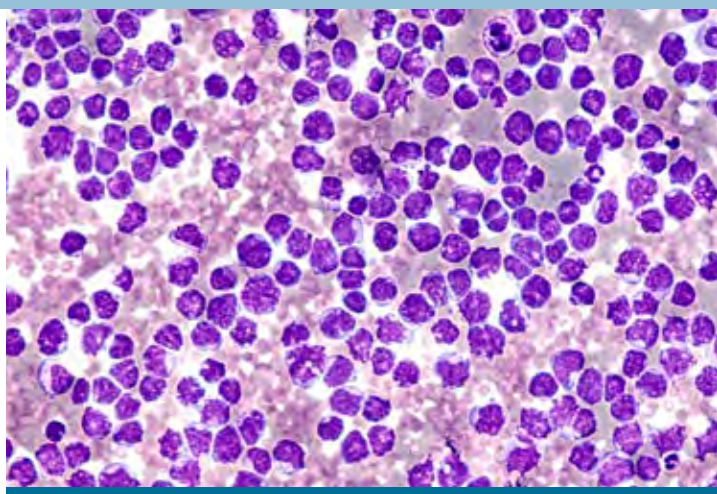
exclusively for a specific cell type, the professional antigen-presenting cells. These are part of the immune defense. Their most important task is to activate the T cells by taking up fragments of pathogens or pathologically altered cells and then presenting them on their surface. This activates the T cells, and the cellular immune response, in which the T cells directly destroy the antigen-presenting cells, is initiated.

The mutant stem cells also present fragments of their own proteins via this mechanism, activating the T cells. These remove the mutant stem cells so that they can no longer serve as blueprints for daughter cells. This elegant protective mechanism actively prevents the development of blood cancer.

New methodology enables precise understanding of leukemia development

The researchers discovered the previously unknown protective mechanism by using the latest technologies. Their real research goal is to understand how leukemic stem cells develop - stem cells that can trigger leukemia due to a genetic change. “In order to cure leukemias, these leukemic stem cells must be specifically killed. So far, this has not been possible. The development of such therapeutic approaches is hampered, among other things, by the fact that too little is known about these blood stem cells,” adds Dr. Simon Raffel, a physician and leukemia researcher at Heidelberg University Hospital and also a member of the LeukoSyStem consortium.

The research work in the consortium is intended to close this knowledge gap – and thus pave the way for new diagnostic and therapeutic approaches, for example for acute myeloid leukemia (AML), a malignant disease of the hematopoietic system in the bone marrow. To study both healthy stem cells and their development into diseased leukemic stem cells, the researchers are developing high-precision methods - so-called single-cell technologies. “Using these methods, we can read out gene activity, surface molecules as well as genetic changes in thousands of individual stem cells simultaneously,” says Dr. Lars Velten, an associate member in the consortium. The researchers merged the information obtained in this way to create a highly precise map depicting the development of the different types of blood cells. This enables them to better understand the developmental process of the cells, as well as the emergence of leukemia cells. “In order to interpret the complex and huge amounts of data that are generated when using such technologies, we are developing novel



In the advanced stage of blood cancer, the blood is flooded with white blood cells.

Photo credit: Adobe Stock / David A Litman

mathematical and computational methods,” explains Dr. Laleh Haghverdi, physicist at MDC and responsible for the development of bioinformatics methods in the consortium.

Using the specially developed single-cell analyses, the researchers in the consortium are now systematically searching for biomolecules that could be relevant for the treatment and diagnosis of leukemias. Ultimately, the researchers hope to track the evolution of AML and identify resistance mechanisms by which the diseased cells resist therapy.

Further information is available at:

gesundheitsforschung-bmbf.de/de/systemmedizin-9458.php

BMBF-Newsletter:

An overview of the most important news from the BMBF in recent weeks (published monthly).
bmbf.de/newsletter

The BMBF keeps you up to date via Mastodon, Twitter, Facebook and Instagram:

social.bund.de/explore

twitter.com/BMBF_Bund

facebook.com/bmbf.de

instagram.com/bmbf.bund

"deep learning enables new approaches in image data analysis"

Interview with Katarzyna Bozek

Katarzyna Bozek heads one of the seven junior research alliances funded by the BMBF's e:Med funding concept. Together with her team, she is using new Deep Learning approaches to investigate multidimensional image datasets of tumor samples to gain new insights into the biology of triple-negative breast cancer.

gesundhyte.de: *Why are triple-negative breast tumors so relevant for research?*

Katarzyna Bozek: Triple-negative breast tumors are among the most aggressive breast cancers, and are associated with high recurrence rates and high mortality. About 15-20% of all breast cancers are triple-negative breast tumors.

Unlike in other breast cancers, biomarkers for personalized therapies are not yet available. The main treatment options are still chemotherapy, radiotherapy and surgical removal. However, we cannot yet predict how patients will respond to these options. In our collaborative project, we are therefore looking for potential markers that can be used to predict disease progression and develop new treatment options.

gesundhyte.de: *What data do you have available for your investigations and what methods do you use?*

Katarzyna Bozek: Our network partners in Berlin generate high-quality multidimensional image data sets from various tumor samples. These provide us with important biological and diagnostic information. Using a combination of different deep learning methods, we thus aim to identify the specific features of the tumors in the high-resolution image datasets.

Deep learning enables more efficient approaches in image data analysis, we can now process and analyze these data more precisely and extract quantitative information. In our collaborative research project, we are developing both methods of supervised learning, so-called supervised methods, and methods of unsupervised learning, unsupervised methods.

gesundhyte.de: *What is the difference between supervised and unsupervised learning?*

"In our collaborative research project, we are therefore looking for potential markers that can be used to predict disease progression and develop new therapeutic options."

Computer scientist Katarzyna Bozek works on the development and application of machine learning, including deep learning as well as statistical and algorithmic methods. Together with her team, she is investigating multidimensional image datasets of triple-negative tumor samples using new deep learning approaches in the BMBF-funded research project. New insights into the biology of this extremely aggressive type of breast cancer could enable new therapy options. (Photo: © Fabian Stuertz)

Katarzyna Bozek: Supervised learning methods are used to divide different groups of patients into groups with different disease progression based on their tumor morphology. We can also use supervised segmentation methods to extract certain characteristics, previously defined by us, from image data. So, for example, cell size or the concentration of a fluorescently labeled molecule. However, if we want to investigate whether there are recurring patterns or characteristics in patients in the data set, we use unsupervised learning methods. Here, we don't define any characteristics or patterns in advance and perform exploratory analysis.

In unsupervised learning, we use algorithms to discover structures in the image data that are related to disease characteristics and disease progression. That is, the algorithm independently searches for possible matching patterns in tissue samples from different patients. In the future, our goal is to use not only individual proteins as biomarkers, but also more complex characteristics such as tissue morphology or a composition of different cell types. This would allow us to potentially classify the diverse group of triple-negative tumors in order to find subgroups or commonalities that might enable a specific therapy.

gesundhyte.de: *What successes have you achieved so far?*

Katarzyna Bozek: I am a bit proud that we are now also introducing methods of self-supervised learning into the analysis of image data. These are methods that allow us to assign numerical values to the image information. These values capture the

> "I'm a bit proud that we are now introducing self-supervised learning methods into image data analysis."

most important visual traits of the images without the need for us to manually mark possible image information for data acquisition in advance. The algorithm automatically recognizes the traits and allows for discovery of previously unknown visual biomarkers.

gesundhyte.de: *What would be a real breakthrough in your field of research?*

Katarzyna Bozek: It would be fantastic if we could cluster image data, just like gene expression data. With the methodological approaches of deep learning, it will hopefully be possible in the future to treat image data like omics data. It would be a real breakthrough to be able to analyze sequence data, gene expression data, and also image data in an integrative manner.

gesundhyte.de: *Your collaborative research project consists of several partners. How did you get involved in this project?*

Katarzyna Bozek: I studied in Berlin ten years ago with one of the current project partners, Adrian Granada. We have kept in touch over the years, and when the BMBF's funding program was published he asked me if I wanted to be involved in the project. Of course, I didn't hesitate for long. We know each other well and that makes communication easier. Good collaborations often last longer than the funding period of a project. It really pays to seek and cultivate them early on.

More about the project

Triple-negative breast tumors are among the most aggressive breast cancers with high recurrence rates and high mortality. They represent approximately 15-20% of breast cancer cases. Unlike other breast cancers, few targeted therapy options and biomarkers are available for this subtype, so we cannot predict how patients will respond to therapy. The main therapeutic approaches are still chemotherapy, radiotherapy and surgical-operative treatment, with a relatively high recurrence rate and serious side effects. The goal of this research consortium is to contribute to the development of novel patient-tailored therapies for triple-negative breast tumors. To this end, we combine novel quantitative single-cell methods for tumor characterization with artificial intelligence-based data analysis and mathematical modeling to study patient tumor tissue and cellular disease models.

Complementing existing knowledge on the genetic heterogeneity of triple-negative breast tumors, we investigate how these tumors differ in their transcriptome and proteome, but also in the composition of the tumor stroma and immune cell infiltrates. To detect complex patterns of tumor cell response to drugs, we use novel machine learning-based data analysis approaches. We use measurements of protein composition, RNA expression, information on tissue topology, and long-term microscopy experiments at the single cell level to try to find correlations with the clinical course of the patients, which can serve as a basis for establishing biomarkers. In parallel, we also try to better understand mechanistically the biology of tumors behind our observations. Our goal is to use this comprehensive, interdisciplinary approach to identify complex, clinically relevant biomarkers based on a combination of morphological characteristics with the RNA and protein composition of tumor cells and their dynamics during the course of treatment.

Using these biomarkers, we aim to enable a more targeted and effective therapy. We also aim to use this strategy to contribute to a better understanding of the biology of triple-negative breast tumors and their heterogeneity, thereby providing the basis for new therapies.

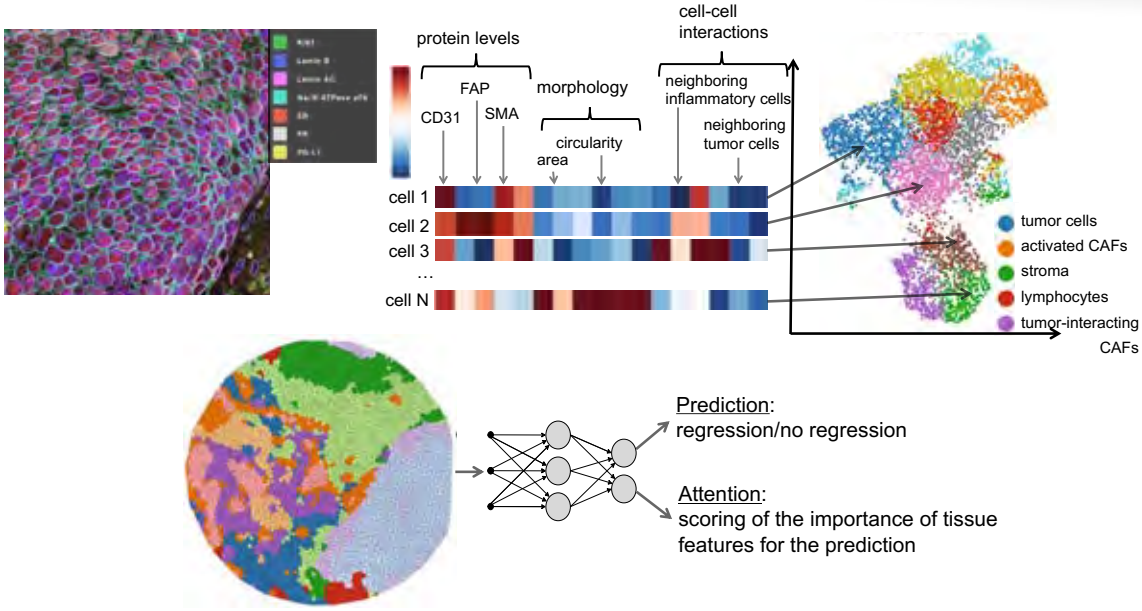


Figure 1: Images play an important role in the project. Multiplex fluorescence images are generated from a large set of patient samples. Using deep learning, relevant information at the single cell level is extracted from the images and used to generate descriptors of the patient samples. These descriptors include proteomic, morphological, and cell type traits. Such a numerical representation of triple-negative breast tumor samples allows not only to search for traits characteristic of patients with certain clinical outcomes, but also to combine the visual with other numerical data such as genome or gene expression (Source: Katarzyna Bozek).

"Good collaborations often last longer than the funding period of a project. It really pays to seek and cultivate them early on."

gesundhyte.de: *What do you hope for in the coming years?*

Katarzyna Bozek: I hope that we will achieve results with our research project that will improve the treatment options for triple-negative breast tumors in the future. But I also hope that my current students will work as scientists anywhere in the world, that they work independently and successfully.

The interview was conducted by Bettina Koblenz and Marco Leuer.

Contact:

Prof. Dr. Katarzyna Bozek
Group Leader Bozek Lab
University of Cologne
Center for Molecular Medicine Cologne
k.bozek@uni-koeln.de

https://www.bozek-lab.cmmc-uni-koeln.de

"I hope that we will achieve results with our research project that will improve the treatment options for triple-negative breast tumors in the future."

preventing common disease with metabolomics and AI

The analysis of metabolomic profiles allows for the prediction of many common diseases simultaneously

by Thore Bürgel

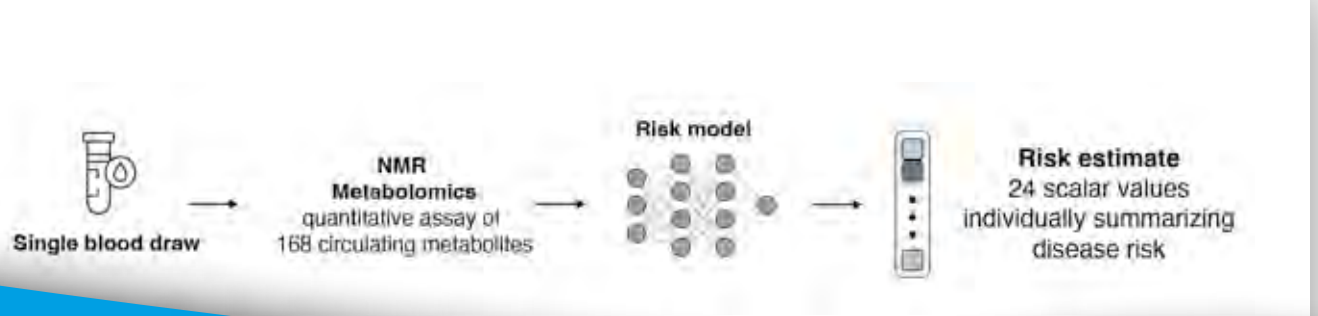
In this study, we investigated the potential of nuclear magnetic resonance (NMR) spectroscopy-derived metabolomics profiles to assess the risk for multiple common diseases at once. We quantified predictive information beyond conventional clinical risk factors for the occurrence of 24 common metabolic, vascular, respiratory, musculoskeletal, and neurological diseases as well as cancers. Our research shows that metabolomic profiles contain important information for risk prediction and could thus contribute to improving the prevention and treatment of common diseases.

Background

In view of the constantly rising costs in the healthcare system, prevention is becoming more and more important as an instrument for reducing suffering and costs. A fundamental prerequisite for the efficient and consistent implementation of preventive measures is the ability to anticipate future diseases as early as possible. Predictive models for risk stratification allow for an individual assessment of disease risks on the basis of measured clinical parameters, an identification of individuals at high risk at an early stage and thus for targeted intervention. Ideally, this happens in the subclinical stage, i.e. before those affected even experience symptoms. However, risk stratification in current clinical practice is limited by a lack of predictive models for most diseases and existing models' limitations to predict single diseases, each requiring specific sets of predictors to calculate. Therefore, the resources (time and cost) required to calculate many disease-specific risk scores hinder the broad clinical acceptance and applicability of risk models in primary care.

Over the past decade increasingly complex information beyond traditional demographics or laboratory measurements have become available in the clinical (research) context, and interestingly, many common diseases are associated with metabolic changes. Metabolomics data provides information about small substances circulating in the blood, such as sugars or fatty acids, that determine human metabolism. This data is thus ought to contain diverse systemic information about the underlying physiology of disease. While some blood metabolites such as cholesterol are already established clinical predictors of cardiovascular disease risk, many other blood

Figure 1: Schematic representation of the approach. Based on a single blood sample and the quantification of 168 blood metabolites, our machine learning model calculates the risk for 24 common diseases simultaneously (source: from Buergel *et al.,* 2022, published under Creative Commons Attribution 4.0 International License, http://creativecommons.org/licenses/by/4.0/).



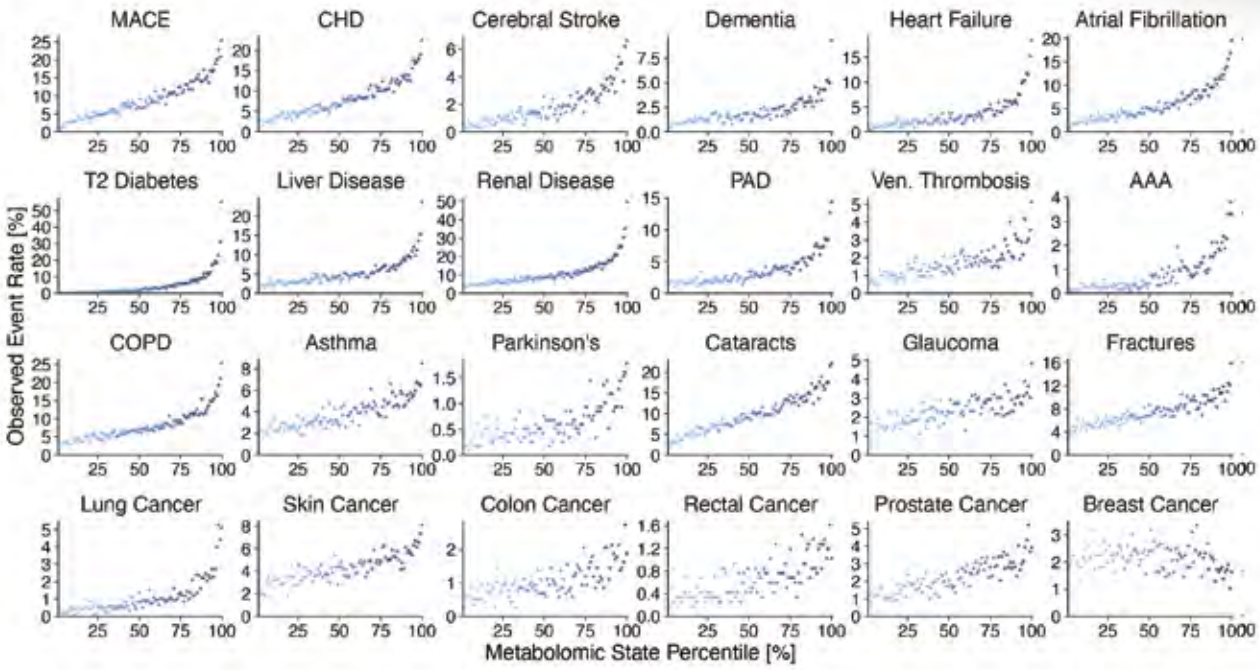


Figure 2: Association with observed disease events. The determined metabolomic state is associated with all assessed endpoints except breast cancer. Event rates per percentile of the metabolomic state are shown. MACE – Major Adverse Cardiac Events, CHD – Coronary Artery Disease, PAD – Peripheral Artery Disease, AAA – Abdominal Aortic Aneurysm, COPD – Chronic Obstructive Pulmonary Disease. (Source: Figure modified by the author from Buergel *et al.*, 2022, published under Creative Commons Attribution 4.0 International License, http://creativecommons.org/licenses/by/4.0/).

metabolites have also been linked to common diseases. Recent work has even gone beyond the association of individual markers and linked entire metabolomics profiles, i.e. the combination of many individual metabolic parameters, to phenotypes such as ageing, disease incidence or mortality. It was found that many diseases have a broad metabolic basis and also often influence the same metabolites by modifying the similar metabolic pathways (Pietzner *et al.*, 2021). The human blood metabolome can therefore be understood as a reflection of the individual's physiological state. This insight led us to the question of whether systemic information from blood metabolomic profiles could be leveraged to simultaneously assess the risk for multiple diseases at once for primary prevention (see Figure 1). Our findings were published in 2022 in the medical journal Nature Medicine (Buergel *et al.*, 2022).

AI allows prediction of many diseases simultaneously

For this study, we relied on NMR-based metabolomic profiles, which allow the simultaneous standardized quantification of a large number of small circulating molecules in the blood. Profiles acquired using nuclear magnetic resonance (NMR) technology differ from comparable technologies such as mass spectrometry in their virtual absence of batch effects, reduced need for expensive reagents and high throughput at comparatively low cost (Würtz et al. 2017; Markley *et al.*, 2017). However, to justify measuring a new data source at primary care sites, a data source must not only be robust and cost-effective, but also contain sufficient additive information beyond predictors already collected, such as age, sex or cholesterol. Therefore, in the present study, we not only quantified the predictive information contained, but also compared it with conventional clinical predictors. We investigated the possibility of using NMR-based metabolomics profiles in a single-sample, multi-disease assay for the simultaneous assessment of 24 common diseases, including metabolic, vascular, respiratory, musculoskeletal and neurological diseases, as well as various cancers. To extract predictive information from metabolomics profiles, we first trained a neural network that summarizes a disease-specific risk from 168 circulating metabolic markers into a scalar value, the *"metabolomic state"*. The model was developed on data from 117 981 participants with a total of ~1.4 million person-years of follow-up from the UK Biobank cohort (Bycroft *et al.*, 2018) and extensively validated on data from four independent cohorts.

Metabolomic profiles contain relevant information for disease prevention

We found that metabolomic states calculated by the model were associated with disease event rates for all diseases studied except breast cancer (see Figure 2). For predictions over a 10-year period, a combination of age, biological sex and metabolomic state was comparably accurate or even better than established cardiovascular predictors for fifteen of the diseases studied. For eight diseases studied, including type 2 diabetes, dementia and heart failure, adding metabolomic state improved prediction beyond very broad clinical variables, suggesting that metabolomic profiles contain information not reflected in conventional clinical parameters.

For other common diseases such as Parkinson's disease, prostate or skin cancer, no improvement over simple demographic predictors such as age and biological sex was detectable. While discrimination, i.e. the accuracy of risk prediction, is the basis for clinical applicability, the clinical utility of a risk model depends on calibration and the choice of appropriate risk thresholds for interventions. Our analysis showed that improvements in prediction resulting from attention to the metabolomic state also translate into clinical benefit for a wide range of endpoints and clinically relevant decision thresholds. Specifically, by considering the metabolomic state, more individuals were correctly categorized at risk for many diseases simultaneously than by risk determination using conventional predictors alone. Beyond these analyses, we identified global disease-associated metabolomic profiles and individual metabolites that were associated with increased or decreased risk of disease onset. For example, in the case of dementia, our approach confirmed numerous established associations of blood metabolites with increased dementia risk, such as fatty acids, including linoleic acid (LA), monounsaturated fatty acids (MUFA) and saturated fatty acids, as well as the protective effect of branched-chain amino acids (BCAA). In addition, our approach also showed evidence of previously unnoticed associations of creatine, albumin and the amino acids glutamine, leucine and tyrosine (see Figure 3).

Pathway to application

In summary, our results show that the acquisition of NMR-based metabolomic profiles could be an attractive candidate for a single-domain multi-disease assay to identify individuals at high subclinical risk and then direct them to interventions or further diagnostic screening. Many countries already recommend regular screening with blood tests for adults for prevention and early detection of selected common diseases, such as the Health Check in Germany (www.bundesgesundheitsministerium.de/checkup.html). Thus, NMR metabolomic profiling could be used in combination with easily collectible clinical predictors such as age and biological sex as a substitute for comprehensive laboratory measurements. In addition, it could conceivably be used to refine risk profiles in combination with comprehensive laboratory testing. While the study population considered was healthier and more privileged than the general UK population, the results of external validation in four independent cohorts suggest general transferability of the calculated metabolomic states. While the NMR-based assay used is currently limited in the breadth of metabolites analyzed and focused on lipids, future expansion of the coverage of NMR-based measurement techniques should further increase the utility. However, before NMR metabolomics can be considered for routine care, many logistical issues, such as sample processing or transport, need to be addressed in further studies.

Overall, our work illustrates the potential of NMR metabolomic profiling for simultaneous assessment of many disease risks to reduce costs, optimize workflows and enable personalized prevention of many common diseases.

References:

Buergel T, Steinfeldt J, Ruyoga G, *et al.,:* Metabolomic profiles predict individual multidisease outcomes. Nat Med 28, 2309–2320 (2022). https://doi.org/10.1038/s41591-022-01980-3

Bycroft C, Freeman C, Petkova D *et al.* The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018). https://doi.org/10.1038/s41586-018-0579-z

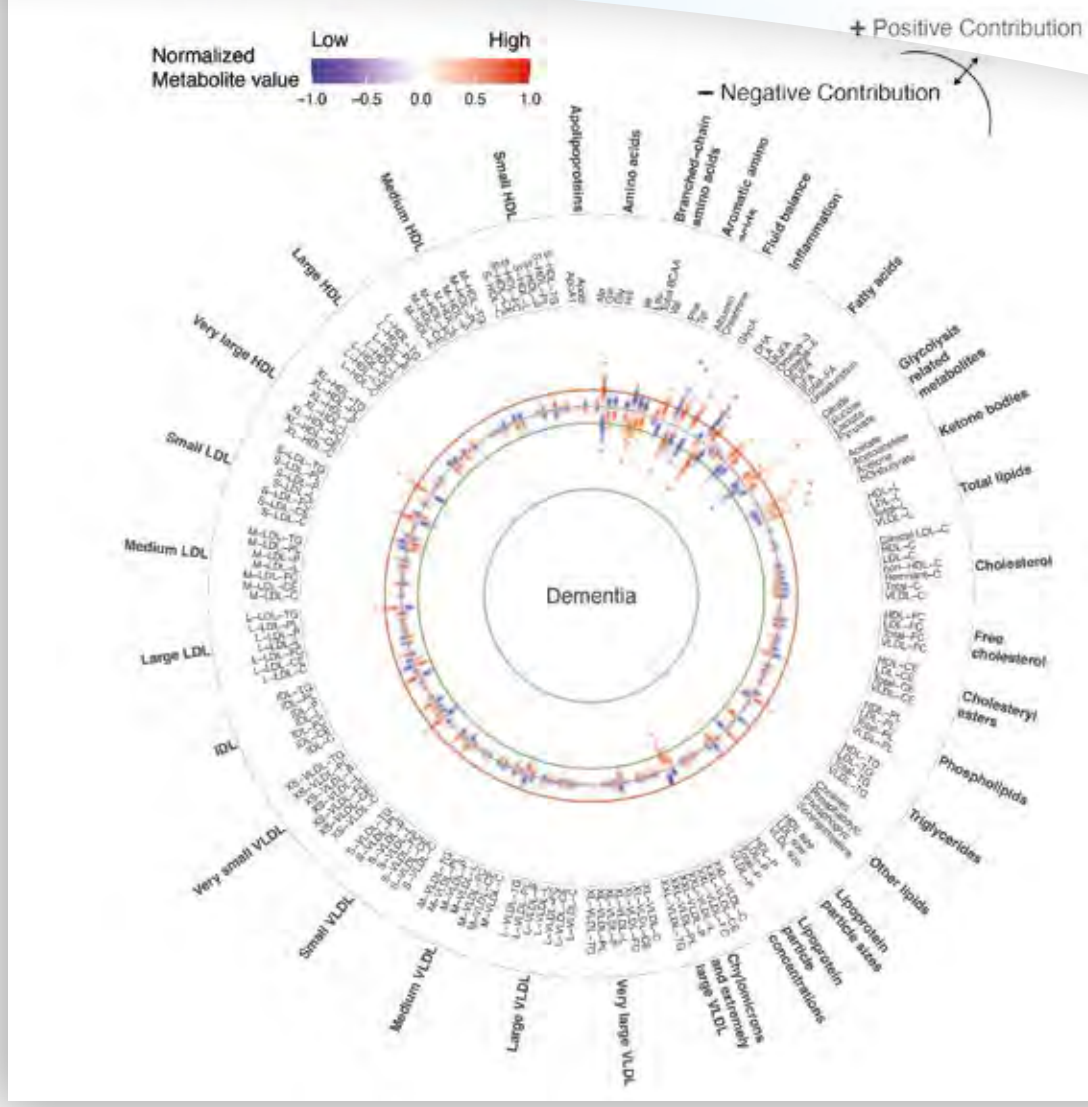


Figure 3: Risk-associated metabolites for dementia. In the diagram, the individual associations are divided into percentiles and each point represents a percentile. The further a point is from the circular baseline, the stronger the absolute association for the percentile. Deviations towards the center represent negative deviations towards the outside represent positive contributions to the metabolomic state. The color indicates the mean measured value of the metabolite in the blood. (Source: Figure modified by the author from Buergel *et al.*, 2022, published under Creative Commons Attribution 4.0 International License, http://creativecommons.org/licenses/by/4.0/).

Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, Wishart DS. The future of NMR-based metabolomics. Curr Opin Biotechnol. 2017 Feb;43:34-40. https://doi.org/10.1016/j.copbio.2016.08.001. Epub 2016 Aug 28. PMID: 27580257; PMCID: PMC5305426.

Pietzner M, Stewart ID, Raffler J, Khaw KT, Michelotti GA, Kastenmüller G, Wareham NJ, Langenberg C. Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. Nat Med. 2021 Mar;27(3):471-479. https://doi.org/10.1038/s41591-021-01266-0. Epub 2021 Mar 11. PMID: 33707775; PMCID: PMC8127079.

Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. Am J Epidemiol. 2017 Nov 1;186(9):1084-1096. https://doi.org/10.1093/aje/kwx016. PMID: 29106475; PMCID: PMC5860146.

Contact:

Thore Bürgel
Center of Digital Health
Berlin Institute of Health at Charité (BIH)
Berlin
tbuergel@pm.me

Photo: Thore Bürgel, private

protein structure prediction in the AlphaFold age

Machine learning enables “structural biology for all”

by Filip König, Karel van der Weg and Holger Gohlke

In order to understand and influence biological processes, knowledge of the detailed spatial structure of proteins is of great importance. Over the last decades, a wide range of different methods for protein structure prediction has been established. Recent developments in protein structure prediction using machine learning provide accurate insights without the need for sophisticated experimental support. This article provides an overview of the different approaches to protein structure prediction and the programs developed in our group for complementary prediction of protein-related properties.

A large proportion of molecular processes in cells are carried out by proteins. Examples include the catalysis of metabolic reactions by enzymes, signal transduction and communication by receptors, and molecular defense against infection by antibodies. In order to enable such diverse processes to be carried out by proteins, a wide range of different protein folds and structural motifs have evolved in the course of evolution. Knowledge of the spatial structure of a protein is necessary to understand in detail how it functions in a biological context. For example, to block the activity of a viral enzyme by rational design of an inhibitor, an accurate spatial description of the active site is essential. A variety of experimental methods are available for the structural elucidation of proteins; among these, X-ray crystallography, nuclear magnetic resonance spectroscopy, and cryo-electron microscopy represent the methods with the best possible resolutions. However, these methods are complex, expensive and without guarantee of success. Therefore, computer-based prediction has emerged in recent years as an alternative way to obtain a protein structure.

Protein Structure Prediction

In order to predict protein structures, a wide variety of approaches have been developed over the past decades. One traditional method is molecular dynamics (MD) simulations. Here, the energy landscape of a protein is sampled using simplified potentials. While MD simulations play an important role in the description of protein structural dynamics, they are only suitable for simulating the folding process of small proteins without further contextual information. On the one hand, the usual time scales over which protein folds are formed are difficult to cover due to the computational power required, and on the other hand, the intrinsic imprecision of the potentials plays a role.

As a promising alternative to protein structure prediction based on a physical description and exploiting structural evolutionary information, the concept of homology modeling has been established quite early. Here, evolutionary conservation of structure is used: If two proteins have a similar amino acid sequence, they usually also have a similar fold. Accurate models can then be generated by comparison with the sequences of all structurally resolved proteins in the Protein Data Bank (PDB). However, the approach is limited if no homologous structures are available in the database. The program TopModel developed in our group is based on the homology modeling approach and uses machine learning approaches on top of that.

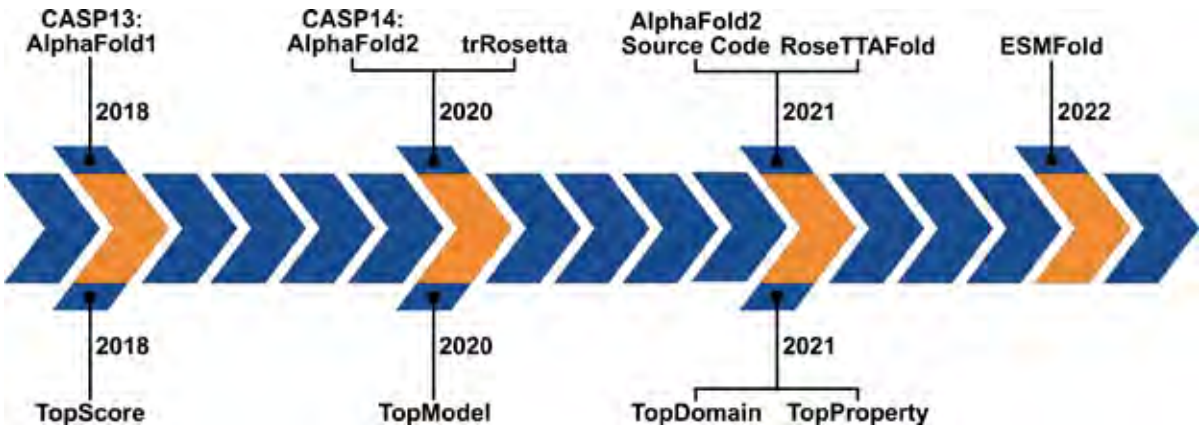


Figure 1: Timeline for the recent development of protein structure prediction programs (Graphic: Filip König).

To predict structures in the absence of resolved homologous structures, coevolution-based prediction has evolved in the last decade. Here, the target sequence is compared with all known protein sequences from databases. Since there are significantly more known sequences than structures, the chance of finding homologous sequences is greatly increased. If concerted mutations are found for a pair of amino acids in the homologous sequences, this can be interpreted as an indication of the spatial proximity of the two amino acids in the folded protein. The basic idea here can be seen in evolution: If an amino acid mutates very frequently together with another amino acid, it can be assumed that their interaction with each other is essential for the structural stability of the protein. This can be interpreted as an indicator of an interaction within the protein. If enough of these pairwise interactions can be found, they can be used as distance constraints between amino acids and thus accurate models can be generated.

In the recent past, prediction models based on deep neural networks (DNN) have gained acceptance. Here, mostly explicitly calculated metrics such as the score of coupled mutations or the retention of amino acids during evolution are used as properties, and architectures such as convolutional neural networks are applied to predict spatial proximity of amino acids within a protein. These contacts are then used in the form of distance constraints, similar to the co-evolution-based approach, to search the thus constrained search space for native conformations. Examples of this type of deep learning-based program include AlphaFold1 and trRosetta (Yang *et al.*, 2020).

The AlphaFold revolution

A major breakthrough in protein structure prediction was achieved by the Google subsidiary DeepMind with its AlphaFold2 program (Jumper *et al.*, 2021). As part of the biennial CASP competition, in which various research groups attempt to predict structures that have not yet been published, AlphaFold2 was able to make predictions for a majority of the target structures that are within the accuracy range of experimental methods. Compared to previous deep learning-based methods, AlphaFold2 uses a whole range of innovations that in combination lead to higher accuracy. For example, AlphaFold2 does not compute explicit coevolution scores, but leaves it to the neural network architecture to extract information about the interactions of amino acids within the protein from multiple sequence alignments. Another key innovation is the end-to-end concept: instead of predicting amino acid contacts and subsequently using them as distance constraints, AlphaFold2 directly provides a protein structure with Cartesian coordinates. In analogy to physical simulation, AlphaFold2 reuses predicted structures as input, resulting in iterative improvement. Currently, more than 200 million structures have already been predicted with AlphaFold2 and made publicly available (https://alphafold.ebi.ac.uk/).

Following the publication of the source code of AlphaFold2 and the simultaneous publication of RoseTTAFold in the summer of 2021, adapted versions were quickly developed with specific capabilities to predict protein complexes. One such development is ESMFold, which uses a similar basic structure as AlphaFold2, but applies a so-called protein language model instead of multiple sequence alignments, thus enabling structure predictions in the shortest possible runtime.

TopSuite

In addition to structure prediction, there are a number of other issues related to proteins that can be addressed with deep learning. These include, for example, assessing the quality of predicted structural models, predicting domain boundaries in multidomain proteins, and predicting transmembrane topologies. To address these issues, the TopSuite software suite has been developed in our group since 2017. It contains programs for homology-based structure prediction (TopModel, Mulnaes *et al.*, 2020), structural model quality assessment (TopScore, Mulnaes *et al.,* 2018), domain boundary prediction (TopDomain, Mulnaes *et al.*, 2021), and transmembrane topology prediction (TopProperty). These programs are metamethods, meaning they use external programs to produce primary predictions, combine them into a normalized set of properties, and then use DNNs to make the final predictions. In this way, the benefits of different primary prediction methods can be maximized. All TopSuite programs are available as web servers: : https://cpclab.uni-duesseldorf.de/topsuite/.

TopModel

TopModel is a protein structure prediction program based on the concept of homology modeling. Since the quality of the resulting protein model depends sensitively on the actual evolutionary closeness of the template to the target protein, special emphasis is placed within TopModel on the detection of false-positive homologs. TopModel uses 12 different programs to identify homologous structures ("templates"). Preliminary models are then generated based on the identified templates and an evaluation is performed using TopScore. In a further step, false positive templates are identified and removed using a DNN-based procedure. TopModel uses a combination of clustering, evaluation with TopScore, and structural refinement to generate the final models. In the presence of suitable homologs, TopModel enables highly accurate modeling of protein structures.

To compare TopModel with AlphaFold2, both methods were applied to proteins from a diverse enzyme dataset, and the quality of the resulting models was compared in terms of TopScore (Figure 2). TopScore returns a value between 0 and 1, with a lower value indicating higher model quality. It can be seen that especially in the lower TopScore range, AlphaFold2 produces models that are rated as more accurate by TopScore compared to the models that TopModel produces. On the other hand, in the higher TopScore range, TopModel produces models that have a better TopScore than those produced by AlphaFold2.

Overall, machine learning-based methods have produced considerable progress in protein structure prediction, and depending on the question, targeted programs can provide further accuracy benefits.

Research project in short form

Name:

InCelluloProtStruct – Hybrid approach to predict the supertertiary and quaternary structure of proteins and protein complexes in cells.

Funding measure:

"Computational Life Sciences" of the Federal Ministry of Education and Research (BMBF)

Participating partners:

Working group Prof. Dr. Holger Gohlke, Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf

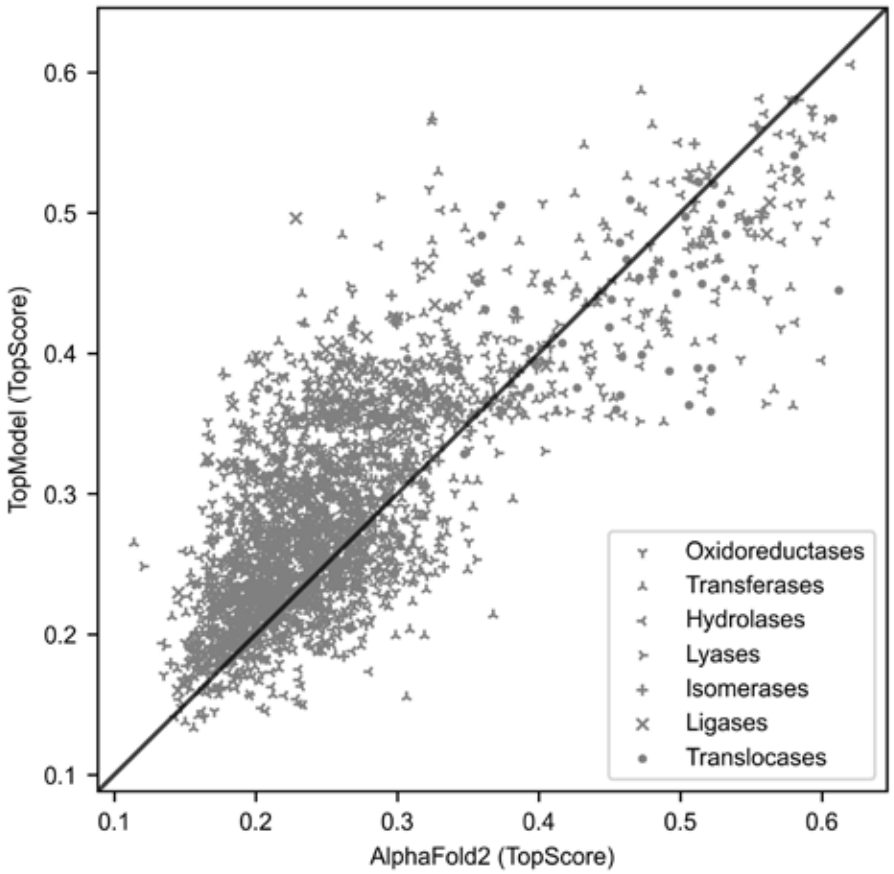


Figure 2: Comparison of TopScore values for structures predicted with AlphaFold2 and TopModel.
Figure adapted based on: https://www.biorxiv.org/content/10.1101/2022.06.13.495871v1
(Graphic: Karel van der Weg).

References:

Jumper, J. *et al.,* (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589.

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S. und Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. Proc. Natl. Acad. Sci. 117, 1496-1503.

Mulnaes, D., Porta, N., Clemens, R., Apanasenko, I., Reiners, J., Gremer, L., Neudecker, P., Smits, H.J.S. und Gohlke, H.. (2020). TopModel: template-based protein structure prediction at low sequence identity using top-down consensus and deep neural networks. J. Chem. Theory Comput. 16, 1953-1967.

Mulnaes, D. und Gohlke, H. (2018). TopScore: using deep neural networks and large diverse data sets for accurate protein model quality assessment. J. Chem. Theory. Comput. 14, 6117-6126.

Mulnaes, D., Golchin, P., Koenig, F. und Gohlke, H. (2021). Top-domain: Exhaustive protein domain boundary metaprediction combining multisource information and deep learning. J. Chem. Theory Comput. 17, 4599-4613.

Contact:

Prof. Dr. Holger Gohlke
Heinrich Heine University Düsseldorf
Institute for Pharmaceutical and Medicinal Chemistry
Düsseldorf
gohlke@uni-duesseldorf.de

https://cpclab.uni-duesseldorf.de

Photo: Steffen Köhler, HHU Düsseldorf

droplet diet for the liver

A look inside the liver cell reveals new molecular players and mechanisms that regulate lipid droplet size – revealing the fundamental pathology of fatty liver disease

by Nachiket Vartak

Pan-European epidemiological surveys show that nearly half of the European population including children suffer from accumulation of fat in the liver – a disease unimaginatively known as "Metabolic associated Fatty liver disease" (MAFLD)[1]. It is in this context, that the BMBF funded the Liver Systems Medicine consortium of scientists and doctors to bring together their efforts to understand and develop treatments for MAFLD. Within this consortium, I established the "Molecular Steatosis and Imaging' group" to understand the very core of MAFLD – the abnormal accumulation of fat in the cells of the liver, using advanced microscopic imaging methods and molecular biology. Through a five-year long effort, asking this basic question about the cell's machinery pointed the way to a potential therapy for MAFLD.

A Curse of the Plentiful

Starve for even a few hours and your body will give you signs of hunger. And yet, the body almost never gives you any indication that you have "too much" food. The roots of this insensitivity to over-nutrition probably lie in our evolutionary history – humans evolved in conditions where starvation was more common and more threatening to survival. Our bodies tend to hoard excess sugars and fat when it is available. Today, plentiful nutrition is available to us in our diet and this tendency to hoard fat turns against us. Modern society is replete with diseases of over-nutrition such as diabetes, obesity and MAFLD. Of these, MAFLD poses the most significant risk due to its widespread nature in the population.

Due to the resilience of the liver, liver disease progresses over many years without any symptoms indicating that something is wrong. By the time symptoms finally appear, patients have already deteriorated to severe complications, often leaving liver transplantation as the last resort.

Conspicuous Signs and Cellular tags

At the cellular level, however, things had already started to go wrong a long time ago. Inside the cell, functional compartments called "vesicular organelles" are chambers where specific reactions and storage can occur. Lipid droplets are such a vesicular organelle, in which fats are stored. The diverse types of vesicular organelles are identified by proteins of the Rab family, which coat their surface. At least 25 different Rab members are present in human cells. Each type of organelle has a specific type of Rab protein on its surface. Lipid droplets, the organelle in which fat is stored, is decorated with the protein tag Rab18 on its surface. Liver cells absorb fat from the blood and store it into lipid droplets. Consequently, the lipid drop-

[1] In 2023, the disease MAFLD discussed in this article was renamed to Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) by a global consensus of scientists and physicians at the European Association for the Study of the Liver (EASL) Congress in order to more accurately reflect the underlying aetiology of these diseases. (Rinella *et.al.*)

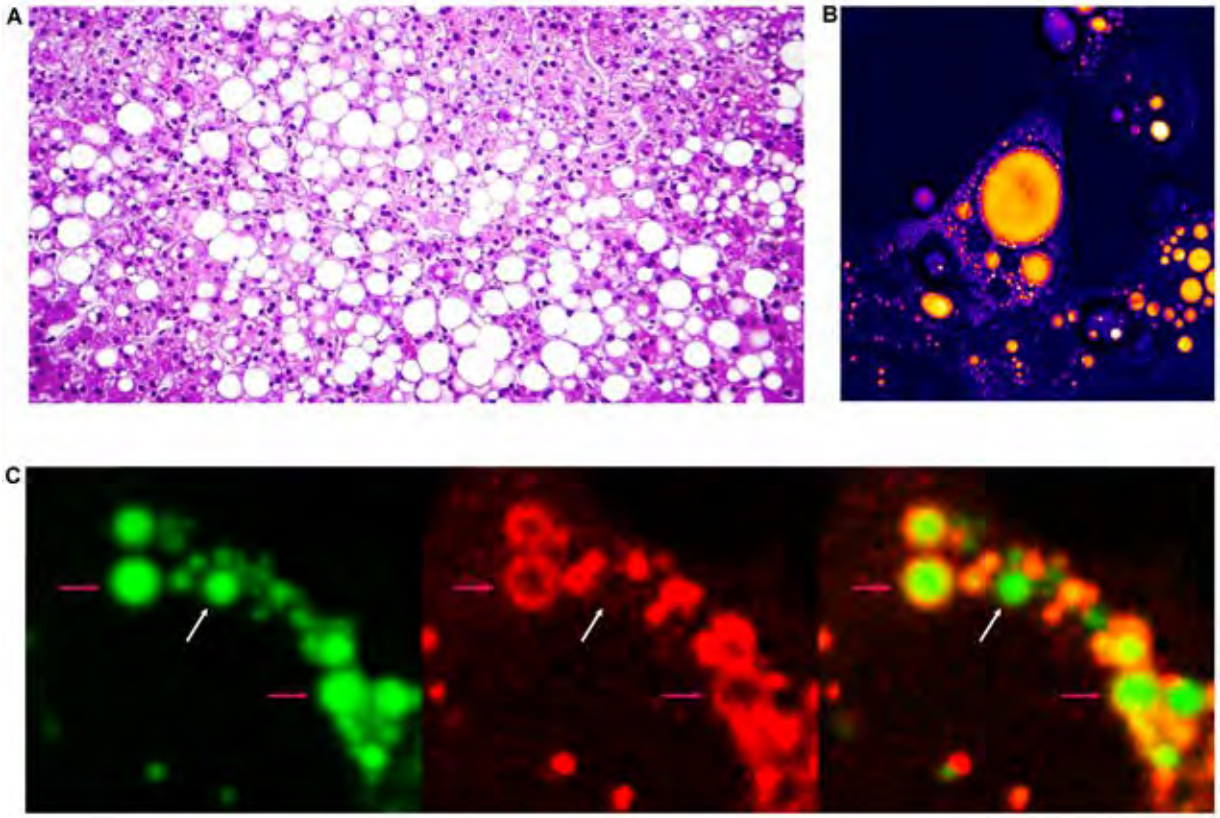


Figure 1: Visualizing fatty liver disease at microscopic scales.
A. In conventional pathology samples, fatty liver disease is characterized by large lipid droplets, appearing as large white empty blobs in liver tissue. Liver cells appear pink.
B. A CARS image of lipid droplets in a liver cell, appearing as intense yellow blobs in a false-color
C. Below: A confocal laser scanning image of lipid droplets (green) and Rab18 tag (red) expressed in cultured liver cells. Droplets indicated with magenta arrows are decorated with Rab18, while the white arrow shows the occasional vesicle with no Rab18 tag on it.

lets grow from initially small vesicles (less than 1µm) to larger vesicles (~2-5 µm). This fat is then utilized as required by the cell by exporting it out of the lipid droplet. When the fat is no longer required, the lipid droplet fuses with another type of vesicular organelle – the lysosome. The lysosome is identified by its own unique Rab tag – Rab7. The lysosome is the "stomach" of the cell – filled with digestive enzymes and high acidity. As the lipid droplet and lysosome fuse into a single vesicle, the lipids are destroyed by the acidic digestive environment of the lysosome.

The most conspicuous sign of MAFLD is characterized by lipid droplets being stuffed with excess fat until their size increases up to 10 times beyond normal. At this abnormal size (~20 µm), they take up almost all the space in the cell, disrupting nearly all the other microscopic structures and cellular functions. These are the cellular foundations of MAFLD.

Peering into the Lipid Logistics

We hypothesized that something went wrong with the logistical control of lipid droplets in MAFLD, and that the Rab18 tag was somehow involved in these lipid droplets becoming very large, forming the foundation of the project. A research team and technical toolset needed to be established to investigate this hypothesis.

To observe lipid droplets without disturbing the normal functions of the cell, a microscope using a special imaging technique called Coherent Anti-stokes Raman Spectroscopy (CARS) was established that directly probes fats, revealing lipid droplets in exquisite detail. To observe the Rab18 tag directly on

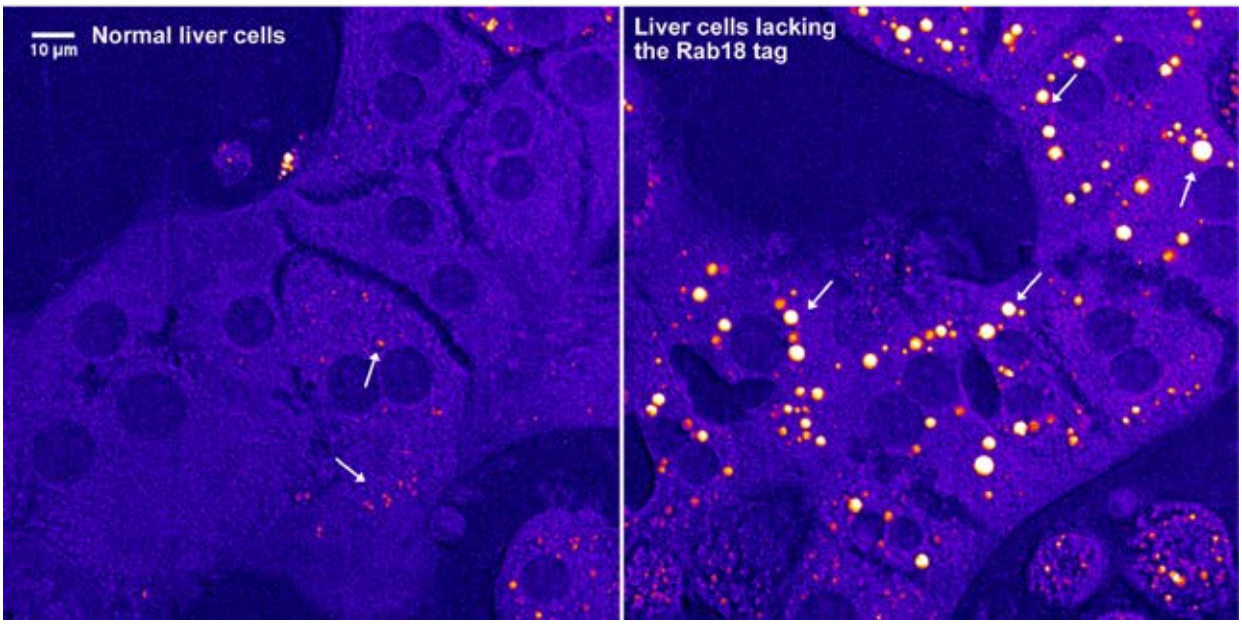


Figure 2: Lipid droplets, shown with white arrow, seen in normal liver cells and liver cells lacking Rab18. The enlargement of lipid droplets in cells lacking is clearly seen in these CARS images. (Source: Nachiket Vartak, published under Creative Commons Attribution 4.0 International License, http://creativecommons.org/licenses/by/4.0/).

the droplets, Adrian Rieck, a doctoral candidate in the group, took a DNA copy of Rab18 gene from humans and fused it with the DNA of a protein called FusionRed that comes from the bubble-tip anemone. This "fluorescently fused" Rab18 glows red under green laser light when seen through an imaging technique called confocal laser-scanning microscopy (CLSM). We also made a variety of "mutants" – versions of Rab18 that are "broken" in precise ways that prevented them from sticking to the lipid droplets.

The processing of microscopy images was automated using a machine-learning algorithm to identify and quantify thousands of lipid droplets in microscope images. This automation allowed measurement of lipid droplet distributions with unprecedented statistical robustness.

Cellular Steatosis

With all the tools now ready, we performed RNA Interference – in which liver cells are prevented from making the Rab18 protein. The cells are then provided with a diet of fats and visualized under the microscope. In normal liver cells which could synthesize Rab18, the lipid was packed into lipid droplets as usual, with the cells making new lipid droplets to store the excess fat. But in cells that had no Rab18 tag, there were no new droplets. Instead, the few droplets that were present grew larger by 4-fold overnight, similar to what is seen in the human MAFLD disease.

When we added back Rab18 in these cells, they were rescued from having large lipid droplets. But when he instead added back the broken "mutants" of Rab18, the cells could not be rescued from having large lipid droplets. These results were also replicated using drugs that prevented Rab18 decorating lipid droplets. Whenever there was no Rab18 on the droplets, the droplets grew large to abnormal sizes.

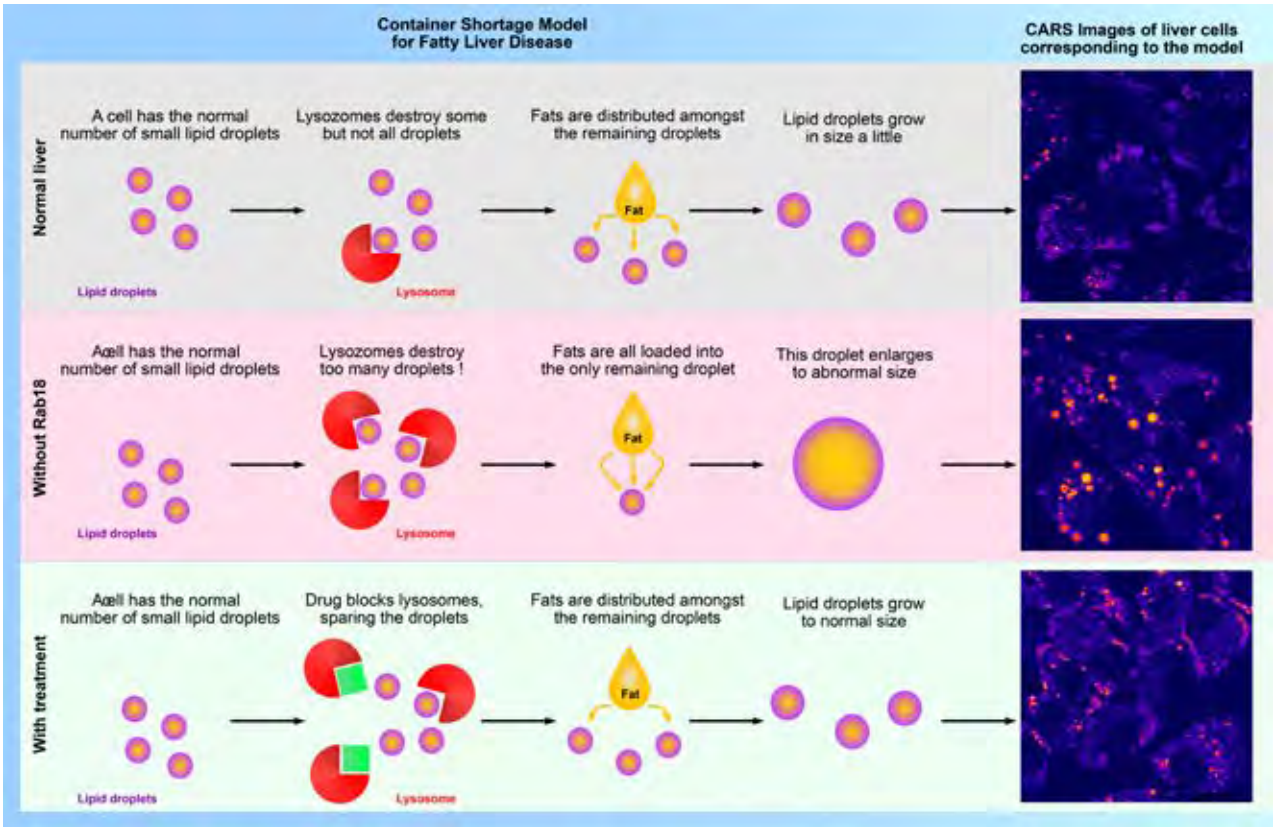


Figure 3: A conceptual model of liver physiology in the context of MAFLD. (Source: Nachiket Vartak, published under Creative Commons Attribution 4.0 International License, http://creativecommons.org/licenses/by/4.0/).

Strangely, removing the very thing that identifies the lipid droplet, leads to the droplets becoming even larger. Should not an unidentified vesicle be ignored or degraded by the cell by fusing it with the lysosome?

Container Shortage

Clearly, a new conceptual model was needed to understand what was going on. So, we went back to the drawing board and came up with a novel hypothesis: Let's say the cell was exposed to excess fat. It already has a few lipid droplets, clearly identified as such with the Rab18 tag. As the cell is given a diet of excess fat, new nascent vesicles are formed and coated with Rab18 to identify them as lipid droplets. The cell now has enough lipid droplets to distribute the excess fat in, and thus each individual droplet has only a small increase in size.

But without the Rab18, the cell could not coat new vesicles with it to designate them as lipid droplets. Instead, these unidentified vesicles would indeed be sent off to the lysosome to be disposed off. There would be a shortage of fat "containers" in cells. All the fat that the cell was being fed would get stuffed into the few lipid droplets it already had. As these few droplets end up absorbing the entire lipid load, they grow to abnormally large sizes!

This "container shortage" model made a testable prediction – if one removed the cells Rab18, there should be many more fusions of nascent vesicles with lysosomes. To test this prediction, the team tracked thousands of lysosomes and newly from nascent lipid droplets in time-lapse imaging of cells, quantifying how many times they interacted when Rab18 was present

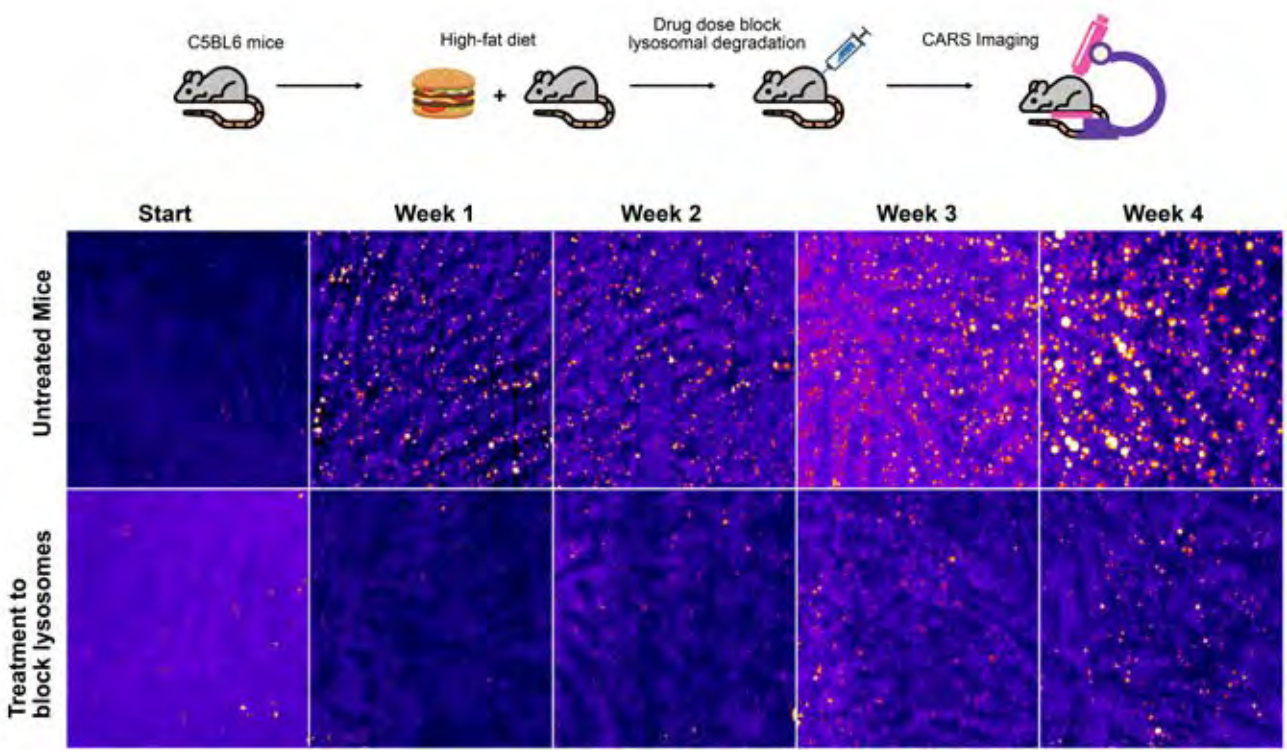


Figure 4: Mice fed on high-fat diet start to develop enlarged lipid droplets, the beginning of fatty liver disease within the first week, getting worse by the fourth week. But mice who recieve a drug to stop lysosomal degradation of lipid droplers show almost no enlarged lipid droplets, showing that the drug protects them from fatty liver disease. (Source: Nachiket Vartak, published under Creative Commons Attribution 4.0 International License, http://creativecommons.org/licenses/by/4.0/).

or not. In cultured cells, or liver cells directly extracted from a mouse and even on human liver cells donated by human patients, the results were the same. Indeed, lipid droplets fused with lysosomes more frequently when there was no Rab18 in the lipid droplets in the cell.

Increased lysosomal digestion of the newly formed small lipid droplets, led to a shortage of lipid droplets in the cell. These then inevitably became abnormally large when they absorbed the entire lipid load. On the other hand, preventing lysosomal degradation to protect the lipid droplets ensured that the cell did not face a container shortage for fats. The lipid droplets then did not grow abnormally large!

Mouse to Man

The final test came from what is considered the closest to the "real situation" that can be reproduced in a lab. Georgia Guenther, an experienced research technician in the MSIM group kept mice on a diet rich in fat for 4 weeks mimicking the overnutrition as seen in human fatty liver disease. But one group of mice were also given a daily dose of a drug that prevented lysosomal degradation. We then performed imaging on the live mice, observing lipid droplets in them with CARS microscopy every week. The mice who had no treatment developed enormous fatty liver disease over the 4 weeks, with lipid droplets growing to twice the size already in the first week.

But mice that were receiving the drug dose did not develop any fatty liver disease at all! Preventing degradation by lysosomes assured that there was no shortage of containers in the livers of this mice.

Thus, asking basic questions about the cell biology of a disease, focused on single protein Rab18 led to clear translational strategy to prevent fatty liver disease that may one day help patients. For reasons of safety, the journey from the lab to the clinic is a long one. In the meanwhile, there is also much more to still discover – for example, if existing MAFLD can be reversed using this strategy or if it is strictly prophylactic. There is scope to improve the drugs that prevent lysosomal degradation by tailoring them for better targeting of the liver.

In either case, inhibition of lysosomal degradation or manipulation of Rab18 localization to the lipid droplet are now novel tools in our arsenal against fatty liver disease.

References:

Pimpin, L. *et. al.,* Burden of liver disease in Europe: Epidemiology and analysis of risk factors to identify prevention policies. J. Hepatol. 69, 718–735 (2018).

Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. Star-convex polyhedra for 3D object detection and segmentation in microscopy. arXiv [cs.CV] (2019).

Rieck, A. The small GTPase RAB18: Insights into cellular steatosis, lipophagy and the non-alcoholic fatty liver disease. (2021) doi:10.17877/DE290R-22423.

Rinella. M. E. *et. al.*, A multisociety Delphi consensus statement on new fatty liver disease nomenclature. Hepatol., DOI: 10.1097/HEP.0000000000000520 (2023).

Contact:

Dr. Nachiket Vartak
Head - Bioimaging Group
Department of Toxicology
Leibniz Institute for Working Environment and Human Factors (IfADo), Dortmund
vartak@bioimaging.tech

Photo: Nachiket Vartak, private

news from the BIH

pioneering method for stem cell diagnostics:

Leif S. Ludwig has received Paul Ehrlich and Ludwig Darmstaedter Early Career Award

Biochemist and physician Dr. Leif S. Ludwig (40) from the Berlin Institute of Health at Charité (BIH) and the Max Delbrück Center has received the 2023 Paul Ehrlich and Ludwig Darmstaedter Prize for Young Researchers. Building on the latest technologies for the gene sequencing of single cells, prizewinner Ludwig has developed a method that can analyse the lifelong regeneration of cells in human blood in a way that is up to 1,000 times quicker, more reliable and less expensive than has previously been possible. In so doing, he is enabling medicine to determine for the first time and with reasonable effort the activity of single blood stem cells in humans.

Our blood renews itself constantly. Each second, millions of new cells are added to our bloodstream which replace dying blood cells. They originate from haematopoietic (blood-forming) stem cells in the bone marrow and then gradually mature over several stages. A distinction is traditionally made between four major developmental trajectories: the first trajectory produces the red blood cells that transport oxygen, the second supplies the thrombocytes, or platelets, that stop bleeding and allow wounds to heal. In the third trajectory, the white blood cells develop, which give us our innate immune defence, such as the granulocytes, for example, and in the fourth, the B and T cells develop, which form the basis for our acquired immune defence in the event of infection. However, as research progressed, the more and more difficult it became to distinguish these trajectories from each other.

Leif Ludwig concentrated on evidencing natural mutations in the mitochondria of blood cells. These cellular powerhouses have their own, much smaller genome of around 16,600 base pairs. Leif Ludwig combined their analysis with the latest single-cell sequencing technologies (single-cell omics), which enabled him to make statements about the actual health status of the cells under examination at the same time. He and his team have meanwhile refined their method in such a way that they can analyse tens of thousands of cells in bone marrow and blood samples from a patient. Leif Ludwig's analytical method now makes it possible to disentangle them more easily in order to identify, for example, at which branch point a leukaemia cell develops or a degenerative change occurs.

On the YouTube channel of the Berlin Institute of Health at Charité (BIH), physician Dr. Leif S. Ludwig answered questions in the video format 3 plus 2 to coincide with the award ceremony. Photo: © BIH/Thomas Rafalzyk

The modern six-story research building will provide 14,875 square meters of floor space for research groups, technology platforms, study centers and outpatient areas, and is directly linked to Charité's high-rise ward block. Photo: © BIH

innovative research and clinical care under one roof:

Rahel Hirsch Center for Translational Medicine of the BIH and Charité opens its doors

On January 19, 2023, the Rahel Hirsch Center for Translational Medicine officially opened next to Charité's high-rise ward block. The opening ceremony was attended by the Governing Mayor of Berlin, Franziska Giffey; by Berlin's Senator for Higher Education and Research, Health, Long-Term Care and Gender Equality, Ulrike Gote; and by the State Secretary at the Federal Ministry of Education and Research, Judith Pirscher. The building, which formerly housed surgical, intensive care and emergency units, was gutted down to the frame and completely refurbished throughout. It will be jointly used by the Berlin Institute of Health at Charité (BIH) and Charité – Universitätsmedizin Berlin for biomedical research, outpatient care and clinical studies.

The BIH and Charité have been united for just over two years. As Charité's third pillar, the BIH has the mission of advancing translational medicine – a process that involves transferring laboratory discoveries as quickly as possible into clinical solutions that benefit patients. Now a new building is bringing the two partners together spatially: In the Rahel Hirsch Center for Translational Medicine (RHC), physicians from Charité will work with scientists from the BIH under one roof.

BIH's integration into Charité becomes tangible

The Board of Directors of the BIH and Charité thanked the federal and state governments for generously funding the construction project. The former surgery and intensive care wing of Charité was completely renovated and transformed into a modern and spacious research, innovation and patient center. The federal government provided the greater part, or 60 percent, of funding for the approximately €100 million renovation, while the State of Berlin contributed 40 percent of the building costs.

Prof. Christopher Baum, Chair of the BIH Board of Directors and Charité's Chief Translational Research Officer, expressed his deep gratitude to both funding authorities: "We are very pleased that the Rahel Hirsch Center now gives the BIH a true center for medical innovation on Campus Charité Mitte. Our focus here is on advancing data science and genomic medicine. Daily interactions with physicians from Charité's clinical and outpatient departments as well as with patients give our scientists valuable insights for their work. Translational research is put into practice here, and exchange between clinicians and researchers is possible at any time. In this way, the Rahel Hirsch Center will contribute to turning research into health."

A great success! For even more information on the opening of the Rahel Hirsch Center for Translational Medicine (RHC) visit the BIH website. Photo: © Konstantin Börner

With Professor Na an excellently qualified colleague takes over the leadership of this important and joint BIH-Charité program.
Photo: © David Ausserhofer

Professor Il-Kang Na named new director of the BIH Charité Clinician Scientist Program

Prof. Il-Kang Na became the new director of the BIH Charité Clinician Scientist Program (CSP) on January 1, 2023. The BIH Johanna Quandt Professor succeeds Prof. Duška Dragun, who headed the CSP until her untimely death in late 2020. In 2021 and 2022, Prof. Britta Siegmund and Prof. Dominik N. Müller took charge of the program on an interim voluntary basis. The Clinician Scientist Program of the Berlin Institute of Health at Charité (BIH) and Charité – Universitätsmedizin Berlin enables physicians-in-training to conduct research alongside their clinical activities at different stages in their careers.

The BIH Charité Clinician Scientist Program (CSP) has continuously evolved over the past eleven years. Today, in addition to the standard CSP, there are also the Junior, Digital and Advanced Clinician Scientist Programs, thus providing structured, tailored support to young doctors in their career development, both during and after residency training. With currently about 150 active fellows and about 200 alumni, the Berlin-based programs are not only by far the largest of their kind in Germany, but according to the German Research Foundation (DFG) they also set best practice standards for the whole country – especially through their quality assurance measures.

Il-Kang Na, a role-model for aspiring clinician scientists

The CSP's new director holds a BIH Johanna Quandt Professorship and is head of the Immune System Defects and Dysfunctions in Tumor Patients Group at the BIH and head of a research group at the Experimental and Clinical Research Center (ECRC). She is also a senior physician in the Department of Hematology, Oncology and Tumor Immunology at Charité Campus Virchow-Klinikum, under the direction of Prof. Lars Bullinger. As a long-time member of the BIH Charité Clinician Scientist Board and as spokesperson of the Berlin School of Integrative Oncology (BSIO), she has additionally gained a wide range of experience in cultivating young medical talent. She wants to further develop the CSP in the longer term and adapt measures that support young physicians in academic medicine to the current situation. "The digital transformation and medical progress on the one hand, and the acute shortage of specialists and the pandemic on the other, pose special challenges for young physicians working in academic medicine today," says Na. "Anyone who wants to seriously pursue research alongside clinical work needs support, especially in our current times, and that's what we want to provide."

TEF-Health will contribute to Digital Europe's overall aim.
Photo: © Petra Ritter/BIH

€60 million committed to establish AI and robotics in healthcare

The EU project TEF-Health aims to test and validate innovative artificial intelligence (AI) and robotics solutions for the healthcare sector and accelerate their path to market. It is led by Prof. Petra Ritter, who heads the Brain Simulation Section at the Berlin Institute of Health at Charité (BIH) and at the Department of Neurology and Experimental Neurology of Charité – Universitätsmedizin Berlin. The 51 participating project partners from nine European countries will receive funding to the tune of about €60 million , with half coming from the European Commission under its Digital Europe program and half from national funding agencies. Some €2 million of the EC funding will go to the BIH.

Technological advances in the field of AI and robotics are being made at a breathtaking pace – and the healthcare sector is not spared from these developments. Yet it goes without saying that new medical devices and procedures must first prove their safety and usefulness before they can be adopted in clinical practice. In the European Union, the areas of AI and robotics, which are set to have a far-reaching impact on the healthcare sector, especially have to meet high quality requirements, but there is still a lack of testing infrastructure for developing standards, validating innovations, and certifying new products.

This is precisely where the Testing and Experimentation Facility for Health AI and Robotics (TEF-Health) comes in. The new project, supported by the EC and national funding agencies with a total of about €60 million, aims to "facilitate and accelerate the validation and certification of AI and robotics in medical devices," explains Prof. Petra Ritter, who coordinates the consortium and heads the Brain Simulation Section at the Berlin Institute of Health at Charité (BIH). Some €2 million of the EU funding will go directly to the BIH at Charité. In total, 51 academic and private partners from nine European countries are involved in the project, integrating existing infrastructures as well as building new ones. The project officially kicks off on 1 January 2023.

Putting new technology through the pace

"With TEF Health we mainly want to test novel AI approaches in real-world scenarios," says Ritter. This includes new software used in areas such as patient care and diagnostics, as well as devices controlled by artificially intelligent programs, some of which are designed for direct use on humans – such as surgical and nursing robots. "We will evaluate how to facilitate market access and acceptance of these smart technologies," Ritter reports.

Photos: © Thomas Rafalzyk

Contact:

Katharina Kalhoff and
Ole Kamm
Press, Online and Social Media Editor
for Science Communication
Berlin Institute of Health at Charité (BIH)
pressestelle-bih@bih-charite.de

www.bihealth.org

EpiBlok

develops gene therapy against focal epilepsy

Company profile EpiBlok Therapeutics GmbH

by Stefanie Seltmann

In May 2022, scientists from the Charité and the **Medical University of Innsbruck** founded EpiBlok Therapeutics GmbH, which is developing a gene therapy for epilepsy patients. Adeno-associated viruses are to deliver the gene for the neuropeptide dynorphin specifically to neurons in the affected brain region. The goal is to suppress seizures in the long term by having the neurons produce dynorphin in stock and release it when needed. Charité BIH Innovation, the joint technology transfer arm of Charité – Universitätsmedizin Berlin and the Berlin Institute of Health at Charité (BIH), supported the founders in patenting the underlying invention.

Neurologists call epilepsy a "thunderstorm in the head": whole groups of nerve cells suddenly discharge, causing unwanted movements or disturbances in the way people feel. About five percent of the population have such a seizure at least once in their lives. In focal epilepsies, where the seizure is triggered in a specific area of the brain, medications often fail and have severe side effects that can also impair learning and memory. "Unfortunately, we can't really help many of those affected very well," says Professor Regine Heilbronn, head of the Gene Therapy working group at the Department of Neurology at Charité – Universitätsmedizin and co-founder of EpiBlok GmbH. "Even surgical intervention does not guarantee permanent freedom from seizures. That's why we developed a completely new therapeutic approach."

"Drug-on-Demand Therapy"

The new approach is based on a so-called gene vector. This contains the genetic information for protective neuropeptides and is introduced directly into the focus of the epilepsy. In focal epilepsy, the production of a small protein, the neuropeptide dynorphin, is often too low. With the help of a gene vector the scientists, introduced the dynorphin gene into the affected nerve cells. These then began to produce and store dynorphin peptides. Professor Christoph Schwarzer, neuropharmacologist at the Medical University of Innsbruck and co-founder of EpiBlok explains what is special about the therapy: "It is a drug-on-demand therapy: the nerve cells only release the stored peptide when it is needed. This is the case when the nerve cells are in constant excitation, as at the onset of an epileptic seizure. Dynorphin inhibits the spread of excitation, and seizure development subsides."

The scientists have already shown in mice that the gene therapy is safe and that it reliably suppresses epileptic seizures for several months after a single application. As a vehicle for the dynorphin gene they use adeno-associated virus (AAV) vectors which are already clinically approved for gene therapy of several diseases. Regine Heilbronn has received 3.3 million euros from the GO Bio funding program of the German Federal Ministry of Education and Research for preclinical studies on the AAV-based gene therapy for focal epilepsy.

Professor Christopher Baum, Chairman of the Board of Directors of BIH and Director of the Translational Research Unit of Charité, also welcomes the spin-off: "In order to make new developments in gene therapy available to patients with previously inadequately treatable diseases, we absolutely need the great drive of scientists who are also taking the challenge of founding a company. In Berlin, we are currently establishing an ecosystem for gene and cell therapies that will further im-

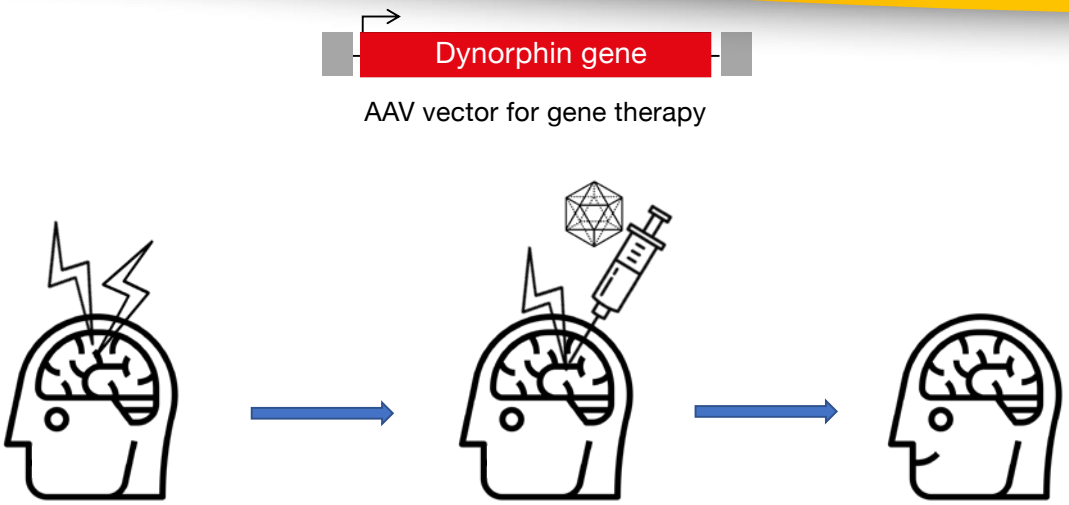


Figure 1: Neuropeptide-based gene therapy for focal epilepsy (Graphic: Design by Regine Heilbronn and Sara Lazaro Petri from Charité and EpiBlok; Syringe by Sergey Demushkin; Pain by Anna Witt; Brain by monkik from the Noun Project).

prove the start-up opportunities for founders. EpiBlok is being established in the right place at the right time. This is how research can turn into health."

Next step: clinical trial

With the help of the newly founded company, the team around Regine Heilbronn and Christoph Schwarzer now wants to take the leap into the clinic. "With EpiBlok Therapeutics GmbH we want to produce the AAV vector in larger quantities and in the required high quality in order to get a first clinical trial off the ground." The SPARK-BIH program supported Regine Heilbronn's team in establishing EpiBlok GmbH. Dr. Tanja Rosenmund, head of the SPARK-BIH program, is pleased with the joint success: "The founding team is developing the first gene therapy which was supported by the Berlin SPARK-BIH program with funding, coaching, mentoring and networking. The goal of SPARK-BIH is to support life science inventions so that more new products and therapies benefit patients. We are pleased that this highly innovative project will now be further developed at EpiBlok GmbH."

Charité BIH Innovation's Patents and Licensing team entered into an exclusive license agreement with EpiBlok in April 2022 for the use of the company founders' invention, for which Charité has filed a patent application. Dr. Bettina Büttner, Technology Manager of the Patents and Licensing team at Charite BIH Innovation, comments: "Patents and the acquisition of an exclusive right of use are an important basis for a spin-off, especially if it is aiming for cost-intensive drug development. This can block imitators and secure the exclusive marketing right. EpiBlok is the first spin-off and the first license partner of Charité to pursue a gene therapy treatment approach. We are very excited about the further development and progress in this exciting treatment option for focal epilepsies."

Find more information on the SPARK-BIH program here:
www.spark-bih.de

Find more information about EpiBlok here:
https://www.bihealth.org/en/notices/epiblok-is-developing-a-gene-therapy-for-epilepsy

Contact:

Prof. Dr. Regine Heilbronn
Co-Founder & CEO
EpiBlok Therapeutics GmbH
regine.heilbronn@charite.de

Photo: © Regine Heilbronn, private

"networking is the model of the future"

Interview with Maike Sander, Max Delbrück Center

gesundhyte.de: *Prof. Sander, what drew you back to Germany from San Diego?*

Maike Sander: What I find very appealing is the potential we possess to collaborate with partners and shape something truly impactful. I view this as a tremendous opportunity for research in Germany and more widely throughout in Europe. Coming from a U.S. perspective, what has struck me in recent years is how interconnected research is here. People strive to work through networks, to address major scientific challenges through joint structures. This is exactly the model we need for the future! Today, it is impossible for a single professor or a single laboratory to bring in or acquire all the expertise necessary to make real progress in biomedicine. In a collaboration, each partner can bring in their unique expertise from across many different disciplines.

In the USA, this way of working is sometimes more difficult, because funding is primarily oriented toward individual research groups, often for limited periods of time. Institutions in the U.S. are not funded by the government per se. Individual professors generally finance their laboratories through third-party funding. Contrast that with the situation here, where research centers such as the Max Delbrück Center can join forces with other institutes, from within the Helmholtz Association and beyond, to tackle the really big questions in health research. I also see great opportunities particularly in the Berlin scientific ecosystem, where we have been making innovative advances in this direction. Look, for instance, at the collaborations between Charité – Universitätsmedizin Berlin, the Berlin Institute of Health at Charité (BIH) and the Max Delbrück Center. With these partners, we have a real chance to shape the medicine of the future. It is the combination of all these factors that made Berlin so appealing for me.

gesundhyte.de: *How efficient is the German research community when you compare it with the major centers in the USA?*

Maike Sander: In Boston, in the Bay Area, or in San Diego, you find very dynamic markets that give the U.S. a clear advantage in many respects. This creates a lively mixture of private investors, venture capital and often huge private donations that drive certain research fields and generate fresh ideas. It constantly creates new models that promote innovation. Here in Germany things are a bit slower; the capital needed to finance start-ups in a new field does not flow quickly or without considerable bureaucracy. On the other hand, in Germany there is more stability and continuity, and that, too, is very important for innovative research.

gesundhyte.de: *What would be necessary to accelerate the momentum?*

Maike Sander: One idea would be to conceptualize ways to mix public and private funding. There need to be clear incentives to promote commercialization so that results are translated into useful diagnostics and therapies quickly. We have

"Today, it is impossible for a single professor or a single laboratory to bring in or acquire all the expertise necessary to make real progress in biomedicine. Collaborative science is what we need."

Prof. Maike Sander is heading the Max Delbrück Center for Molecular Medicine in the Helmholtz Association in Berlin as Scientific Director since November 2022. She is an internationally active stem cell and diabetes researcher. Sander became known for her groundbreaking research, particularly on insulin-producing beta cells, whose development and function she analyzes with her team.
(Photo: © Peter Himsel / Max Delbrück Center)

very solid governmental funding in Germany, excellent scientists, and great ideas. But in order for us to bring these innovations into practice, this module is still missing: Start-ups need dynamic options for private funding, as well as a large capital market.

gesundhyte.de: *How can this be achieved?*

Maike Sander: Again, it's all about networks. In San Diego, we had events several times a year where investors came and scouted for opportunities. Such platforms are an ideal environment for an ongoing exchange of ideas; professors and investors in the U.S. enjoy very close ties. And when ideas emerge, the path to realization is very short. So, the big question is: how can we promote such a culture and attract investors over here? How can we show them that our ideas and discoveries will eventually be very profitable? I see that as a task for us as a network: Together, we can show what Berlin has to offer and promote interest in the location. This will create more meetings and networks, especially including investors.

THE POTENTIAL OF TALENT

gesundhyte.de: *How do you see the Max Delbrück Center and the life sciences developing in the capital region?*

Maike Sander: The enormous potential of Berlin starts with the fact that we have three major universities that are already cooperating as the Berlin University Alliance. Add to that all the non-university research institutions, which have an enormous appeal. These have also joined forces as the "Berlin Research 50." This represents a significant strength of the city: a level of density of scientific activity that is not often encountered. Many of the crucial building blocks we need for good

"Together, we can show what Berlin has to offer and promote interest in the location. This will create more meetings and networks, especially including investors."

"Berlin has tens of thousands of students, an incredible potential source of talent! Beyond that, ideas, diversity, creativity and innovation have to be lived, which is another advantage of Berlin, because it is so attractive as a metropolis."

science are already on hand. That goes beyond great research; Berlin has tens of thousands of students, an incredible potential source of talent! Beyond that, ideas, diversity, creativity and innovation have to be lived, which is another advantage of Berlin, because it is so attractive as a metropolis. This means we can attract talent and professionals from all over the world. Within our triangle formed by Charité, BIH and the Max Delbrück Center, I see our role as a driver of innovation. We are on the forefront of new technologies, and we are furthering their development by applying them to the most complex questions related to human diseases. These advances are happening in close cooperation with clinicians. Now we need to consider what strengths and focal points already exist, and how we can expand them and further establish ourselves as global leaders.

gesundhyte.de: *What specifically do you have in mind?*

Maike Sander: "From bench to bedside and back" is a principle we are already living at the Biotech Campus in Berlin-Buch, where the main Max Delbrück Center campus is located. I am

Biography Prof. Maike Sander

Prof. Maike Sander is an internationally active stem cell and diabetes researcher. Since November 2022 she is heading the Max Delbrück Center for Molecular Medicine in the Helmholtz Association (Max Delbrück Center) in Berlin as Scientific Director.

Sander studied medicine at the University of Heidelberg and then spent more than 25 years conducting research in California. Most recently she served as director of the Pediatric Diabetes Research Center at the University of California in San Diego. Sander became known for her groundbreaking diabetes research, particularly on insulin-producing beta cells, whose development and function she analyzes with her team. The researchers produce pancreatic islets from human pluripotent stem cells. These organoids mimic the natural microenvironment of beta cells in the Petri dish and allow detailed analyses. Regenerating beta cells or replacing them using human pluripotent stem cells could become a therapeutic approach for both forms of diabetes. "It is not yet clear how to transplant beta cells derived from stem cells into people and protect them from rejection in the body," Sander emphasizes. "To find answers to these questions, science and industry must collaborate intensively."

She is a recipient of the Juvenile Diabetes Research Foundation's Grodsky Award and the Alexander von Humboldt Foundation's Research Award. In 2022, she received the Albert Renold Award from the European Society for the Study of Diabetes (EASD).

Discovery for tomorrow's medicine: The Max Delbrück Center is one of the leading international biomedical research centers. Around 1700 people from more than 70 countries work together here in interdisciplinary teams. Their goal is to understand the molecular basis of disease and health and to bring their findings to patients as quickly as possible.
(Photo: © Katharina Bohm / Max Delbrück Center)

talking about the Experimental and Clinical Research Center (ECRC), which we founded together with the Charité in 2007. Another project that follows this principle is our new idea of creating the Berlin Cell Hospital. We want to bring new technologies such as single-cell analysis or CRISPR methods into the clinic. These technologies allow us to obtain an extremely detailed view of disease processes, at a much higher resolution than ever before. This is a precondition to take prevention to the molecular level, to diagnose and treat diseases much earlier, even before patients exhibit overt symptoms and before the body has suffered irreparable damage. This is only possible through close collaborations between clinicians and researchers who are able to collect and interpret large datasets.

gesundhyte.de: *Are there other examples?*

Maike Sander: One fantastic area of progress has been the development of model systems we call organoids. These are mini-organs which we produce in the lab, grown from patients' own stem cells. They serve as new models to study human disease processes. Organoids are becoming increasingly important for research and industry, but to fulfill their potential in practice, we have to learn to automate their production and scale up the process. The aim is to use high-throughput screening methods to determine the point at which disease deviates from health, and whether some active substance can step in right at that point and intercept the pathological process. To achieve this, we will need thousands of high-quality organoids at a time. That would represent an enormous step forward, but it will require that many disciplines must come together, particularly bioengineering and data science. I see an opportunity here that goes well beyond our ambitions at the Max Delbrück Center. The way forward is to expand and develop these opportunities across Berlin, through even closer collaborations.

The interview was conducted by Jutta Kramm and Jana Schlütter.

Contact:

Prof. Dr. Maike Sander
Scientific Director
of the Max Delbrück Center
Berlin
office.sander@mdc-berlin.de

https://www.mdc-berlin.de

with system against addictive disorders

SysMedSUDs: A systems medicine approach to explore resilience and pathomechanisms in substance use disorders

by Lea Zillich, Eric Poisel and Rainer Spanagel

The e:Med project SysMedSUDs funded by the Federal Ministry of Education and Research (BMBF) takes a systems medicine approach, bringing together experts from various fields of medicine and life sciences, to advance research into the fundamentals and possible intervention options for addictive disorders. A special focus is on finding out to what extent there are overlapping mechanisms between substance use disorders and which are substance-specific. In addition to studying samples of patients, we analyze postmortem brain tissue from patients with different substance use disorders, while simultaneously modeling substance use and addiction-related behaviors in animal models and brain organoids, so-called mini-brains.

Why do we study addictive disorders?

Addictive disorders are now classified as substance use disorders ("SUDs"). SUDs are disorders characterized by persistent, uncontrolled, and compulsive use and relapse behavior of legal and / or illegal substances ("drugs"). The dependence syndrome that occurs in SUDs results in the maintenance of substance use despite negative effects on physical and mental health. SUDs are associated not only with the direct consequences of substance use but also with numerous sequelae and comorbidities, thus contributing significantly to the burden of disease on a national scale (1). In addition to negative effects on individual mental and physical health, individuals with SUDs are often affected by stigma and problems in the social environment, which leads to increasing isolation of those affected and can further exacerbate substance use (2). SUDs thus pose a significant challenge to the health and social care system.

Substance use is prevalent in many ways in societies and cultures around the world. According to the World Health Organization (WHO), approximately billion people worldwide drink alcohol, 1.3 billion people smoke tobacco, and nearly 250 million people use illicit drugs[1]. While for a large number of people substance use in the sense of experimental use occurs only once or occasionally as recreational drug use, a not inconsiderable proportion of people develop substance dependence. For example, epidemiological studies in Germany revealed an estimated number of 1.6 million persons with alcohol use disorder and 4.4 million persons with tobacco dependence (3). It should be noted that the risk of developing a substance use disorder varies depending on the substance consumed. The calculated probabilities for the development of dependence were 22.7% for alcohol use, 67.5% for tobacco smoking, and 20.9% for cocaine and 8.9% for cannabis use (4).

These observations raise the question of why certain individuals develop a SUD while substance use in others does not lead to the diagnostic criteria of a dependence syndrome. Previous research suggested that both environmental factors, such as a social environment in which intoxicating substances are regularly used, and genetic risk variants contribute to the risk of developing SUDs (5). The heritability of different SUDs ranges from 0.39 for hallucinogens to 0.72 for cocaine (to be interpreted as 1.00 = fully heritable, 0 = not heritable) (6). This sug-

[1] https://www.who.int/publications/i/item/9789241565639

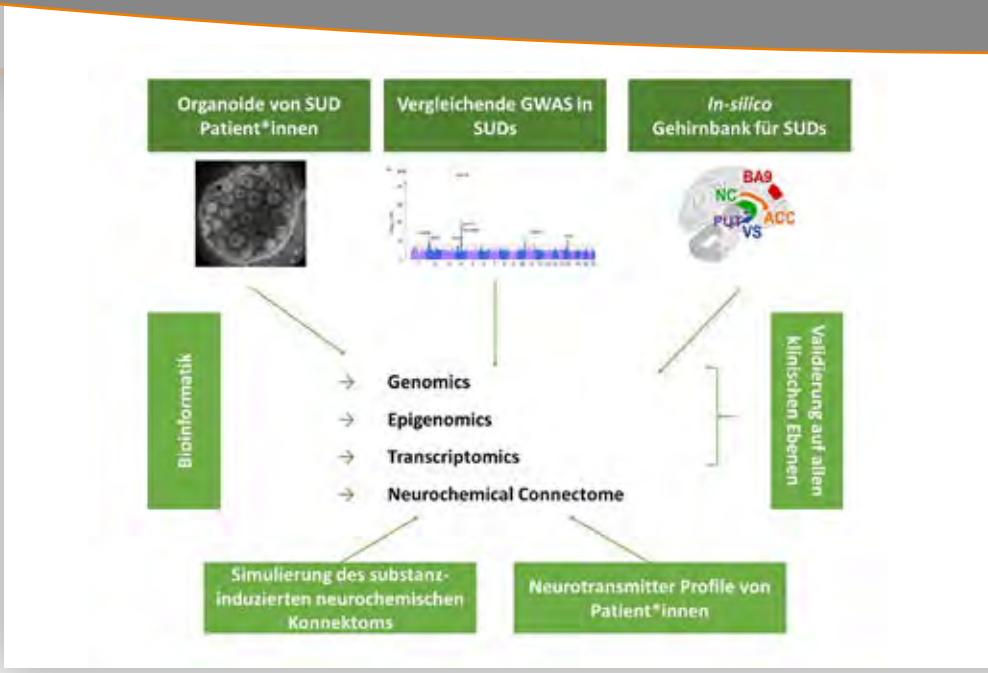


Figure 1: Illustration of the subprojects and how they synergistically interact
Studies on different spatiotemporal system levels. On a genome-wide level genetic variants (GWAS) are deteceted, altered methylation patterns (EWAS), different chromatin states (ATACseq), altered transcriptome states and dynamics (RNAseq) on the cellular level (using the newly developed 10x Genomics Single Cell Multiomics technology) in postmortem brain tissue from SUD patients and controls and also within bulk tissue, mini brains and blood are examined (Source: SysMedSUDs).

gests differential involvement of specific genetic variants that are inherited between generations and thus contribute to a higher risk of disease. In recent years, genetic association studies with large numbers of cases have led to the identification of risk genes that could functionally be located, at least in part, in the mechanism of action and metabolism of the corresponding substances (5).

Another fundamental question in addiction research is how different SUDs differ and what common pathological phenomena exist. Systems medicine approaches represent a promising methodology to increase the understanding of pathomechanisms in SUDs, while offering opportunities to improve diagnostics and treatment options. In addition, findings from studies on SUDs have high sociopolitical relevance in that they can serve as a basis for decisions regarding legalization and taxation of substances.

Neurobiological mechanisms of SUDs

Alterations at molecular, cellular, and systemic levels in the brain are considered to underlie the development of addiction. Repeated substance use induces molecular changes in specific brain regions, particularly in the brain's extended reward system (7). Repeated exposure to the substance leads to changes in epigenetics and gene expression in neurons, which affects communication and interaction between neurons mediated by proteins (7). Thus, under sustained substance use, existing neuronal connections are weakened (e.g., in prefrontal cortex areas leading to a reduction in cognitive flexibility) and new connections between neurons are formed (e.g. in the dorsal striatum leading to habit formation) and strengthened. This process forms the basis for the concept of a so-called "neurocircuitry of addiction" on the basis of which the presence of altered neuronal connections and activities between brain areas in SUDs is postulated. Furthermore, it is hypothesized that substance-induced permanent changes in brain function are responsible for the lifelong behavioral abnormalities in patients with SUDs (8). For example, the presentation of a substance-associated stimulus, such as smoking in a movie, can induce craving for substance use even after prolonged abstinence, contributing to a significant risk of relapse even years later.

The SysMedSUDs Consortium

The goal of the SysMedSUDs Consortium (https://www.sysmed-sud.org) is to identify cross-substance and substance-specific genetic, epigenetic, gene expression, and neurochemical alterations in alcohol, nicotine, heroin, cannabis, and cocaine use disorders (see Figure 1). For this purpose, large cohort samples of SUD patients from different international consortia (e.g. the

Psychiatric Genomics Consortium and the UK Biobank) as well as postmortem brain tissue from deceased SUD patients and corresponding control samples are investigated. This study follows a convergent approach at multiple system levels, such as the genome, epigenome and transcriptome.

At the same time, transcriptional changes in the developing brain are investigated in brain organoids exposed to substances and analyzed comparatively with post-mortem human brain tissue. In further subprojects of the consortium, the neurochemical connectome in SUDs are modeled using modern in-silico methods, and a MR spectroscopic analysis of neurotransmitter profiles in patients with SUDs are performed. To validate and investigate the functional relevance of these findings, animal models are used that show high visual, construct, and predictive validity in SUDs. In the SysMedSUDs consortium, researchers in six subprojects at three different locations (Mannheim, Berlin and Tübingen) contribute to scientific progress in the study of SUDs.

Multi-omics analyses in the SysMedSUDs consortium

A special focus in the SysMedSUDs consortium is on multi-omics analyses, which are characterized by an integrative investigation of data sets from individual omics levels, e.g. the epigenome or the transcriptome. The goal of this methodology is to identify SUD-associated molecular mechanisms at individual omics levels and to summarize the results based on the relationships among different omics levels (see Figure 2).

As part of the SysMedSUDs consortium, we previously performed multi-omics studies that identified differential DNA methylation and gene expression patterns in postmortem brain tissue in alcohol use disorder (9, 10). In particular, similarities between DNA methylation and gene expression in immune processes were evident at the network level (10). The joint analysis of DNA methylation and gene expression represents only one possible example of multi-omics approaches. Upcoming investigations and integrative studies at additional omics levels are ongoing to deepen the understanding of molecular mechanisms in SUDs. Considering the molecular biology dogma from DNA to RNA to protein, other regulatory levels such as small RNAs as well as the final functional level, the proteome, are of particular interest. While it is not yet possible to map all molecular processes in the cell simultaneously, extending analyses to additional omics levels may provide a more detailed picture of the molecular interplay in SUDs. Our experimental efforts to advance multi-omics analyses in addiction will not only provide us with new candidate genes, but also reveal new molecular intervention options.

Outlook

In addition to expanding the levels of analysis in terms of multiomics studies, advances in the field of single cell analysis in recent years have opened up new possibilities in addiction research. For example, RNA sequencing at the single cell level (scRNA-Seq) can be used to analyze the gene expression profile of each individual cell in postmortem brain tissue, which makes it much easier to assign findings to individual cell types due to the heterogeneous composition of the brain (e.g. neurons, glial cells, endothelial cells). Meanwhile, first multi-omics investigations on single cell level are also possible, e.g. parallel analysis of gene expression profile (scRNA-Seq) and chromatin (scATAC-Seq) in the same cell. Within the SysMedSUDs consortium, this opens up the possibility of associating SUD-associated changes in gene expression, e.g. in different types of neurons, with changes in chromatin, thus allowing a more precise characterization of the molecular processes in different SUDs.

Since postmortem brain tissue is a limited resource and modern molecular biological methods allow the production of so-called “mini-brains” – a complex brain organoid can be modeled from easily obtained tissue samples of blood or skin by means of reprogramming in cell culture. The advantage of mini-brains compared to other in vitro models is the ability to model different cell populations, which is very close to the original situation in the brain. In addition, external influences on molecular processes in cell culture experiments can be minimized and systematically excluded by reprogramming experiments. Such mini-brains will then be examined for changes in chromatin and gene expression at the single cell level after acute and long-term substance exposure in a multi-omics approach.

Systems medicine approaches, such as those in the SysMedSUDs consortium, are promising methods to make precision medicine, already established in other fields, accessible for the treatment of SUDs. Given the current paucity of available therapies with high relapse rates and an enormous disease burden, systems medicine for SUDs can be a tremendous benefit to both patients and society.

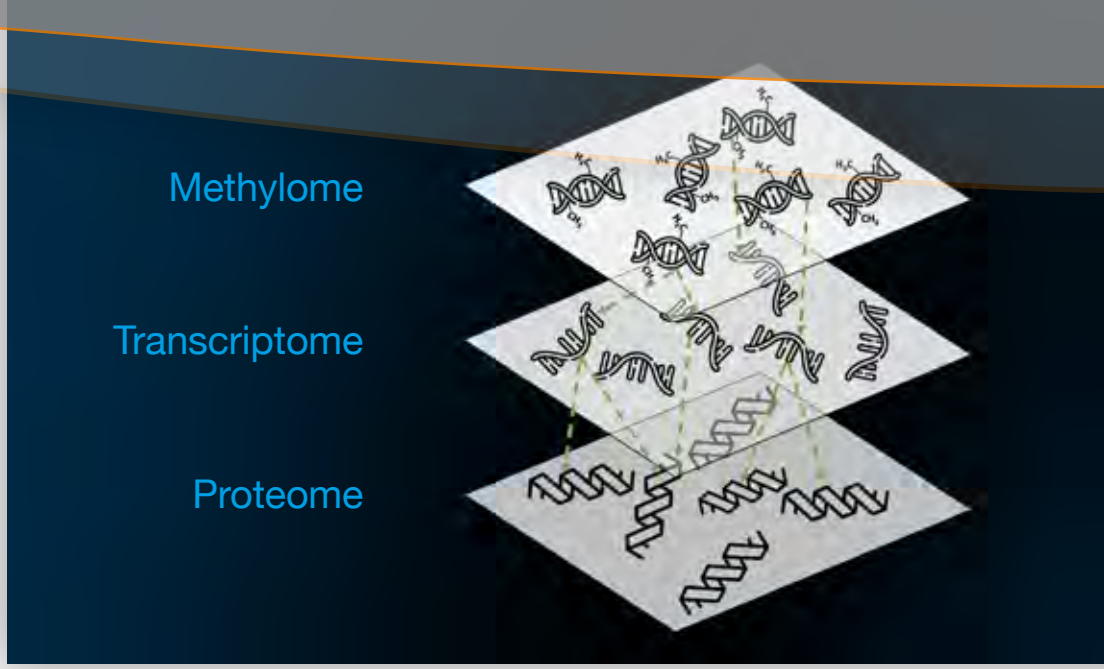


Figure 2: Model representation of a multi-omics analysis of DNA methylation (methylome), gene expression (transcriptome) and protein expression (proteome).
(Source: based on Argelaguet *et al.* (2018)).

References:

1. Effertz T, Mann K. The burden and cost of disorders of the brain in Europe with the inclusion of harmful alcohol use and nicotine addiction. Eur Neuropsychopharmacol. 2013;23(7):742-8.
2. Zwick J, Appleseth H, Arndt S. Stigma: how it affects the substance use disorder patient. Subst Abuse Treat Prev Policy. 2020;15(1):50.
3. Atzendorf J, Rauschert C, Seitz NN, Lochbühler K, Kraus L. The Use of Alcohol, Tobacco, Illegal Drugs and Medicines: An Estimate of Consumption and Substance-Related Disorders in Germany. Dtsch Arztebl Int. 2019;116(35-36):577-84.
4. Lopez-Quintero C, Pérez de los Cobos J, Hasin DS, Okuda M, Wang S, Grant BF, et al. Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Drug Alcohol Depend. 2011;115(1-2):120-30.
5. Deak JD, Johnson EC. Genetics of substance use disorders: a review. Psychol Med. 2021;51(13):2189-200.
6. Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. Nat Rev Genet. 2005;6(7):521-32.
7. Koob GF, Volkow ND. Neurocircuitry of addiction. Neuropsychopharmacology. 2010;35(1):217-38.
8. Hamilton PJ, Nestler EJ. Epigenetics and addiction. Curr Opin Neurobiol. 2019;59:128-36.
9. Zillich L, Frank J, Streit F, Friske MM, Foo JC, Sirignano L, *et al.* Epigenome-wide association study of alcohol use disorder in five brain regions. Neuropsychopharmacology. 2022;47(4):832-9.
10. Zillich L, Poisel E, Frank J, Foo JC, Friske MM, Streit F, *et al.* Multi-omics signatures of alcohol use disorder in the dorsal and ventral striatum. Transl Psychiatry. 2022;12(1):190.

Contact:

Prof. Dr. Rainer Spanagel
Scientific Director
Institute of Psychopharmacology
Central Institute of Mental Health (CIMH)
Mannheim
rainer.spanagel@zi-mannheim.de

www.zi-mannheim.de

Lea Zillich
Scientific Assistant
Department of Geneitc Epidemiology
in Psychiatry
Central Institute of Mental Health (CIMH)
Mannheim
lea.zillich@zi-mannheim.de

Eric Poisel
Scientific Assistant
Department of Geneitc Epidemiology
in Psychiatry
Central Institute of Mental Health (CIMH)
Mannheim
eric.poisel@zi-mannheim.de

machine learning to detect concomitant disorders of schizophrenia

Identifying multimodal comorbidity signatures in psychosis: the e:Med COMMITMENT project

by Emanuel Schwarz and Andreas Meyer-Lindenberg

Patients affected by schizophrenia experience a substantially increased risk for concomitant diseases, including cardiovascular disease and metabolic syndrome. The concomitant diseases (comorbidities) are present in addition to the primary clinical conditions. These comorbidities strongly contribute to the decreased life expectancy associated with schizophrenia, and used to be considered a secondary consequence of medication effects and differences in lifestyle. In fact, one of the core reasons for non-compliance to second-generation antipsychotic medication is the frequent occurrence of metabolic side effects. However, over the last decade, mounting evidence has supported that the biology of cardio-metabolic illness and schizophrenia are intrinsically linked.

Deciphering these shared mechanisms is not easy though, as schizophrenia has its typical onset between adolescence and early adulthood, but affected individuals experience cardiometabolic comorbidities only years, if not decades later. Thus focusing on patients affected by these comorbidities would go hand-in-hand with a relatively long disease duration of schizophrenia. Due to medication, lifestyle and other effects, this would likely cause unfavorable consequences for identifying the shared disease biology. The "COMorbidity Modeling via Integrative Transfer machine learning in MENtal illness (COMMITMENT)" BMBF-funded e:Med alliance tries to address this challenge via an innovative approach from the area of artificial intelligence (Schwarz *et al.*, 2021).

How can machine learning help?

In general, machine learning allows the identification of patterns in data that are predictive of a given (clinical) outcome, such as the schizophrenia diagnosis. In a more advanced approach, the so called "multi-task learning", patterns are learnt that are simultaneously predictive of multiple outcomes, such as schizophrenia and cardiovascular disease (**Figure 1**). COMMITMENT builds on the hypothesis that such patterns capture the shared biology of these conditions. Under this hypothesis, the patterns can be learnt from multiple patient cohorts affected by either condition, without the need for data from patients actually affected by the comorbidity. This has the major advantage that data would be less affected by disease-unrelated effects, and available at far larger scale, a condition that is

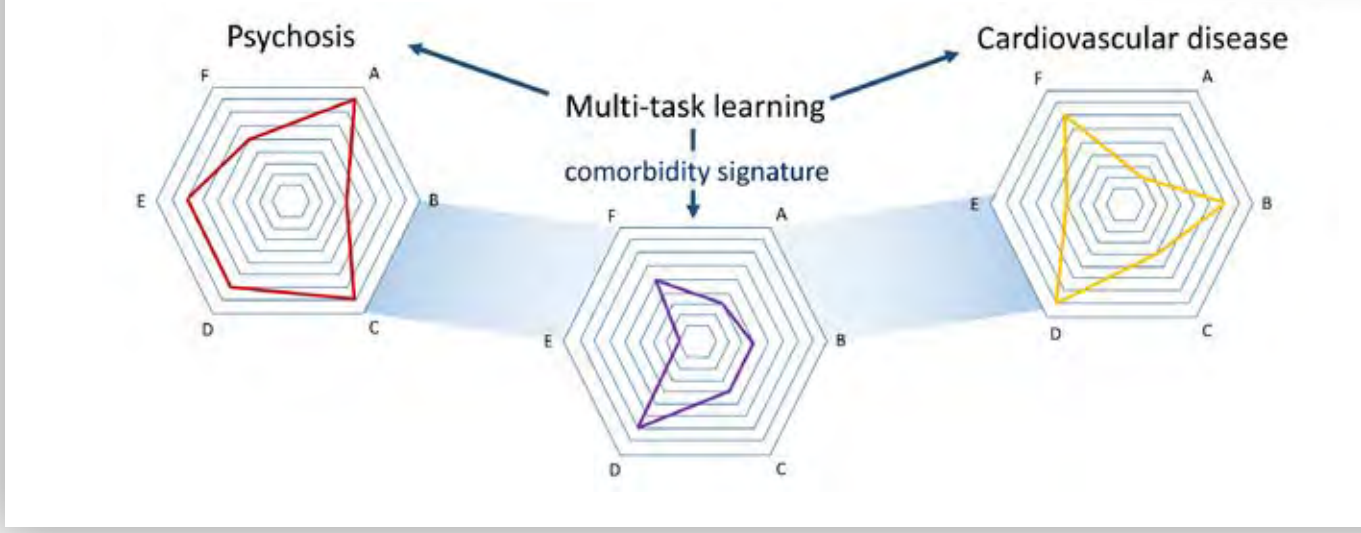


Figure 1. Schematic illustration of how multi-task learning derives a shared, comorbidity-relevant signature from the simultaneous analysis of different patient groups. The comorbidity signature is then validated in patients affected by the comorbidity.
(Graphics: Emanuel Schwarz, ZI Mannheim).

strongly favorable for building predictive machine learning models from high-dimensional data, such as neuroimaging or genetic data.

Thus, COMMITMENT is directly targeted at stratification, the attempt of "separating" patients into subgroups that differ with respect to important biological and clinical properties. The hope is that such differences meaningfully relate to differences in disease mechanisms that provide new insights into how patients should ideally be treated, into the development of new, effective therapies, or prevention strategies. For example, patients with patterns indicative of an illness biology shared with that of metabolic conditions may be at an increased risk of metabolic side effects of certain antipsychotic medications. The machine learning approach pursued in COMMITMENT assumes that such stratification pattern may be "dimensional", meaning that comorbidity-associated differences may not relate to a clearly defined subgroup of affected patients, but show a gradual representation across the entire spectrum from strongly present to entirely absent. This also allows different mechanistic dimensions to co-occur within a given individual and the relative importance of such mechanisms to differ between individuals.

The dimensionality challenge

One of the core challenges of identifying predictive patterns is the high dimensionality of biological datasets, such as those describing inter-individual differences of the genetic code. The challenge arises from the fact that differences associated with clinical phenotypes are typically small and are hidden in a vast sea of changes that are not relevant for the phenotype of interest. This challenge becomes even more pronounced when, as is the case in COMMITMENT, different data types are combined, in order to identify more predictive signatures and gain deeper insight into comorbidity-relevant biological mechanisms. To guide the search for predictive patterns, it is helpful to know in advance which features (such as genes) may be of potential relevance, and this can be informed by the already existing evidence described in the literature. Instead of a manual search, however, COMMITMENT applies automated text mining to extract relationships of likely mechanistic relevance for schizophrenia from the literature. These relationships, which may for example be protein-protein-interactions previously found changed in affected individuals, are then encoded in the form of a "knowledge-graph", which makes them usable by artificial intelligence approaches (**Figure 2**). In a simple implementation, the relationships can be used for the filtering

Photo: AdobeStock © ryzhi

of the biological data, and thus reduce data dimensionality for downstream application of machine learning approaches. In a more advanced implementation, the knowledge graphs are used directly as part of such learning approaches, in order to identify signatures that are more likely to be of mechanistic relevance for a given clinical condition. This can also open up the perspective of querying specific mechanistic hypotheses that are supported by evidence described in the scientific literature. But reducing complexity is not only essential when building artificial intelligence models, it is also critical when translating such models into clinical use. Of core importance is transparency, such that medical practitioners and patients can understand how a given model arrives at a certain prediction. But also from an implementation perspective, simpler models are likely preferable. Not only because they are usually more reproducible, but also because of the increased burden to the patient, as well as the costs, associated with measuring data across multiple platforms.

Focus on the entire life span

The COMMITMENT project has a specific focus on exploring such mechanistic effects across the lifespan. On the one hand, it aims to track how the biomarker patterns identified at the patient group level develop across the life-span. This can aid in differentiating between elements of the signature that are independent of age (the so-called “trait markers”) and those that change with age (the so-called “state markers”). On the other hand, COMMITMENT builds on the concept of the so-called “normative modeling”. This approach is analogous to that of child growth charts, where the behavior of certain variables, such as head circumference, are modeled across the lifespan, using a large reference population. The growth charts then allow to evaluate the “distance” of a given individual with respect to this reference at a particular age. In COMMITMENT, this concept is applied to explore normative trajectories, and deviations thereof, of neural and cognitive measures of importance of psychosis and cardio-metabolic illness. By measuring the deviation of an individual with respect to the expected trajectory, the project aims to shift the analysis focus from the patient group to the individual subject level, and thereby decipher inter-individual heterogeneity in the investigated

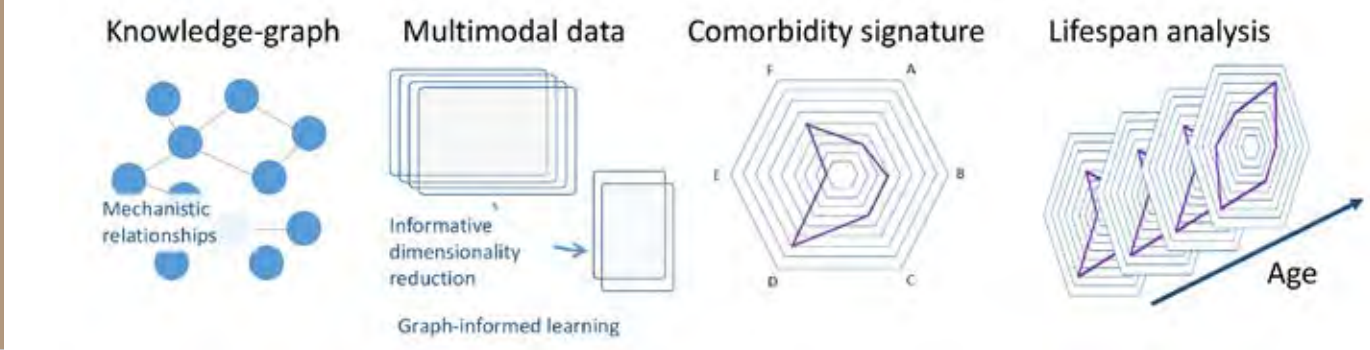


Figure 2. Schematic illustration of COMMITMENT analysis steps. Knowledge graphs inform the meaningful reduction of data dimensionality, and are used directly as part of machine learning approaches, in order to identify comorbidity signatures. These are then evaluated across the lifespan, in order to identify risk periods and characterize trajectories of illness-relevant mechanisms (Graphics: Emanuel Schwarz, ZI Mannheim).

parameters. If the project is successful in identifying trajectories of mechanistic importance to comorbidity, it may allow pinpointing specific age-periods within which relevant biological changes occur. This may, in turn, inform the development of new approaches to treatment and prevention. When this information is integrated in clinical decision making, interventions may range from tapering or switching medication based on (metabolic) side effect profiles, lifestyle measures (movement, diet) up to targeted cardiological procedures. In this way, COMMITMENT hopes to bring down the serious excess mortality of schizophrenia, which costs patients more than 15 years of life on average.

Reference:

Schwarz E. *et al.*, Identifying multimodal signatures underlying the somatic comorbidity of psychosis: the COMMITMENT roadmap. Mol Psychiatry, 2021. 26(3): p. 722-724 https://doi.org/10.1038/s41380-020-00915-z.

Contact:

Dr. Emanuel Schwarz
Clinic for Psychiatry and Psychotherapy
Central Institute of Mental Health
Mannheim Medical Faculty
of the University of Heidelberg
emanuel.schwarz@zi-mannheim.de

Prof. Dr. Andreas Meyer-Lindenberg
Clinic for Psychiatry and Psychotherapy
Central Institute of Mental Health
Mannheim Medical Faculty
of the University of Heidelberg
Andreas.Meyer-Lindenberg@zi-mannheim.de

www.zi-mannheim.de

https://www.sys-med.de/de/verbuende/commitment/

Photos: © Emanuel Schwarz, private; Andreas Meyer-Lindenberg, private

scars on the way – fathoming fibrosis with AI and omics

The synergistic use of AI, single cell sequencing, and spatial proteomics methods show organ fibrosis at an unprecedented resolution – the e:Med Fibromap alliance

by Ivan G. Costa, Rafael Kramann, Rebekka Schneider and Victor Puelles

Tissue fibrosis, or scar formation, is the common final pathway of virtually all chronic diseases and affects nearly every organ, including kidney, heart, lung, and bone marrow. It often starts after damage to the tissue due to a circulatory disorder or an inflammatory reaction.

Well-known examples of organ fibrosis are liver cirrhosis due to excessive alcohol consumption or infections, fibrosis of the heart or kidneys, for example due to high blood pressure, or bone marrow fibrosis due to genetic predisposition. Fibrotic disease represents a large and growing health care burden. However, despite contributing to as much as 45% of deaths in the world, fibrotic diseases have been largely overlooked for many years resulting in a dire therapeutic gap. While it has been accepted among scientists that myofibroblasts are fibrosis-causing cells, their cellular origin and cellular mechanisms associated with their activation is an open field of investigation and remains controversial.

To address these questions the BMBF-funded e:Med Fibromap alliance explores cutting edge molecular technology as single cell sequencing, mouse genetic tracing and spatial proteomics (multiplexing imaging) together with computational models. A major aim is to create spatial maps of fibrotic tissues in order to dissect cellular and molecular interactions associated with myofibroblast activation. For this, the alliance combines the medical expertise of Rafael Kramann (UK Aachen) on kidney and heart fibrosis, of Rebekka Schneider (UK Aachen) on bone marrow fibrosis with the expertise on computational genomics methods of Ivan G. Costa (UK Aachen) and advanced imaging techniques of Victor Puelles (UK Hamburg-Eppendorf).

The computational model CrosstalkeR reveals the Fibrosis-associated signal transmission

Ivan Costa and his team asked the central question: How do the cells involved in fibrosis communicate with each other? They assumed that communication between the different cell types (cellular "crosstalk"), would provide crucial clues to the development of fibrosis. The team therefore investigated the use of ligand-receptor expression analysis from single-cell sequencing experiments to model communication between different cell types of cellular crosstalk. They specifically developed the computational model "CrosstalkeR" for this purpose, which encodes ligand-receptor and cell-cell interactions as networks. It then uses computational concepts such as random walks and network topology measures to find cell types, ligands, or receptors whose influence on a disease increases or decreases (Nagai *et al.*, 2021; Figure 1). In collaboration with Rebekka Schneider, they used this method to analyze cell communication by single-cell sequencing of stromal cells in bone marrow fibrosis (Leimkühler *et al.*, 2021).

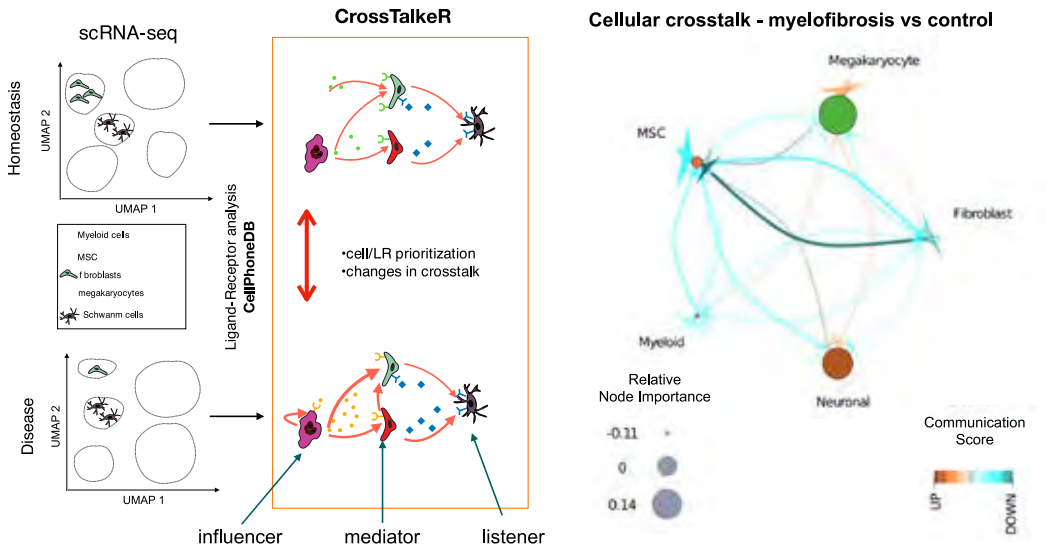


Figure1: Schema on how to obtain cell-cell communication networks from single cell sequencing with CrosstalkeR (left). Analysis of myelofibrosis cells indicates the importance of Megakaryocytes and increase in communication towards stromal cells (MSC) in myelofibrosis (right). Adapted from Nagai *et al.*, 2021.

Indeed, cellular crosstalk increased significantly from dysplastic megakaryocytes to fibrosis-causing mesenchymal stromal cells via TGFb signaling. CrosstalkeR also showed an increase in alarmin (S100A8/S100A9)-mediated signaling from monocytes to mesenchymal stromal cells.

Thus, the researchers discovered that two distinct subsets of mesenchymal stromal cells are pro-fibrotic. These were functionally reprogrammed in a stage-dependent manner and initiated differentiation toward fibrosis formation in the profibrotic stage, whereas they became profibrotic and inflammatory in the fibrotic stage.

The researchers subsequently determined in the mouse model and in patient samples that the aforementioned alarmins are early markers of fibrosis (so-called fibrosis markers). Their therapeutic use of tasquinimod, a small molecule that inhibits these alarmins, actually led to a reduction in bone marrow fibrosis in mouse models (Leimkühler *et al.*, 2021). Currently, Rebekka Schneider is leading a clinical trial to translate these findings to patients suffering from myelofibrosis. With this result, not only a clue to the development of fibrosis has been found, but also directly a starting point for its therapy.

The microenvironment of fibrosis is crucial after myocardial infarction

The replacement of dead cardiac cells with fibroblasts is part of the healing process after myocardial infarction and important to prevent cardiac rupture. However, exaggerated scaring (or fibrosis) is detrimental to heart function. To understand microenvironment and molecular mechanisms associated with heart fibrosis, Rafael Kramann generated a unique resource with single cell sequencing and spatial transcriptomics of human heart specimens after myocardial infarction (Figure 2; Kuppe *et al.*, 2022). Together with Ivan G. Costa, a computational approach was devised to map single cell transcriptomics and epigenomics information into spatial transcriptomic data. This included the inference of a gene regulatory network driving the differentiation of progenitor fibroblasts towards fibrosis causing myofibroblasts (Li *et al.*, 2023). Moreover, the team could characterize a macrophage sub-population, which is particularly co-localized with myofibroblast cells in tissue regions with scarring. The therapeutic modulation of these cells to ameliorate fibrosis is currently under study. Altogether, this indicates how the combination of spatial and single cell techniques can reveal how fibrotic processes are influenced by the tissue microenvironment.

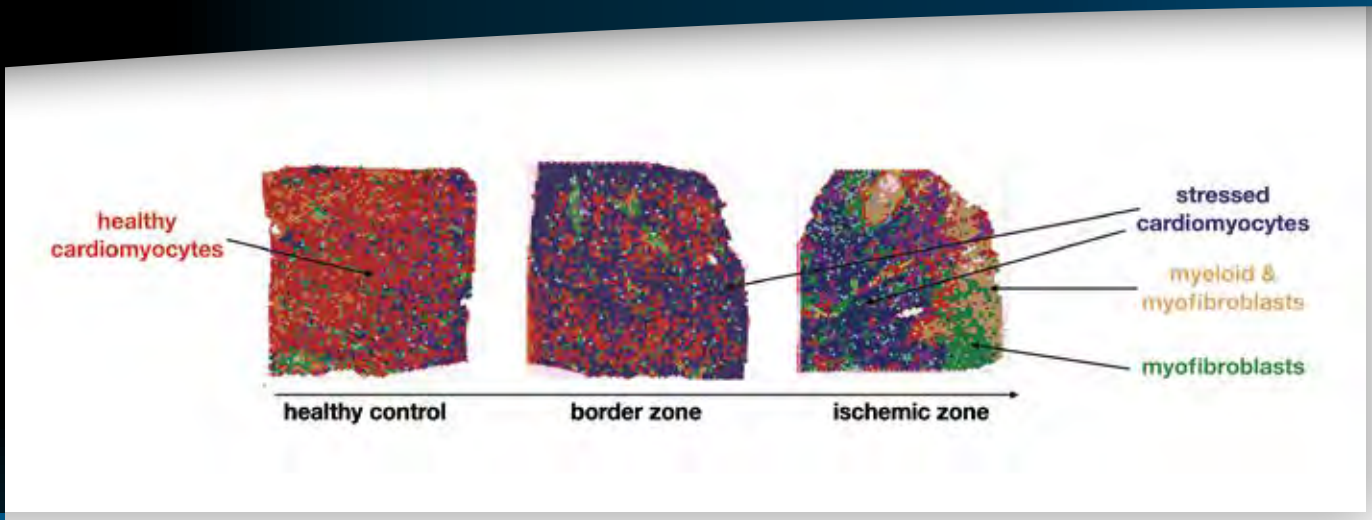


Figure 2: Example of cellular niches found in the process of myocardial infarction. We could detect microenvironments associated with immune invasion and fibrosis initiation. Adapted from Kuppe *et al.*, 2022.

Deep learning-based analysis of kidney biopsies

Fluorescent microscopy is a powerful tool to measure both molecular and morphological features of kidney structures, and it is especially important to detect pathological signs to better understand human disease. In a recent study, Victor Puelles worked together with Rafael Kramann to develop a deep learning-based algorithm to identify podocytes, kidney cells with limited capacity for regeneration that if lost represent a direct cause of kidney filter fibrosis (Zimmermann *et al.*, 2022; Figure 3). For this particular study, we used biopsies of patients diagnosed with Anti-Neutrophilic Cytoplasmic Autoantibody (ANCA) glomerulonephritis, an auto-immune disease characterized by severe damage to the kidney filters, which with time become fibrotic. Our algorithm automatically generated multiple outputs to quantify the degree of podocyte loss, which was used for proof-of-principle experiments for disease diagnosis and prognosis. This study provides a foundation to understand the origins of organ fibrosis at a single cell level in pathological specimens, thereby providing new tools to stratify patients, and personalize therapies.

COVID-19 can induce kidney fibrosis

Data on patients with severed COVID-19 infection indicated frequent loss of kidney function. An important question was if this was a direct effect of the virus, independent of a systemic response due to e.g. infection or blood pressure effects of the infection. Rafael Kramann and Rebekka Schneider explored single cells from kidney organoids infected with SARS-CoV-2 and biopsies from COVID-19 patients by RNA sequencing to characterize them (Jansen *et al.*, 2022). Organoids are organ-like structures a few millimeters in size that are generated using cell culture methods to perform studies on them outside the patient, for example, for functional disorders. Computational analysis by Ivan Costa and his team then showed that in both organoids and patients, COVID-19 infected epithelial cells of the kidney, but not stromal cells. The infected epithelial cells on the organoids activated the fibrosis-inducing TGFbeta signaling pathway, which is not the case with immune cells, confirming direct infection.

In this context, the researchers also observed that a SARS-CoV-2 protease blocker was able to reduce SARS-CoV-2 infection. Imaging experiments performed in collaboration with Victor Puelles also confirmed the presence of increased tubular interstitial renal fibrosis in autopsy specimens from patients. These results demonstrate that SARS-CoV-2 can directly infect renal cells and induce cellular injury leading to fibrosis. They explain both acute kidney injury in COVID-19 patients and the development of chronic kidney disease as a result of long COVID.

Flow cytometric analysis of COVID19 kidney organoids by clinical scientist Dr Katharina Reimer and technical assistant Susanne Schmitz in the laboratory of Professor Rebekka Schneider-Kramann.
Photo: © RWTH Aachen Press/Photo Office

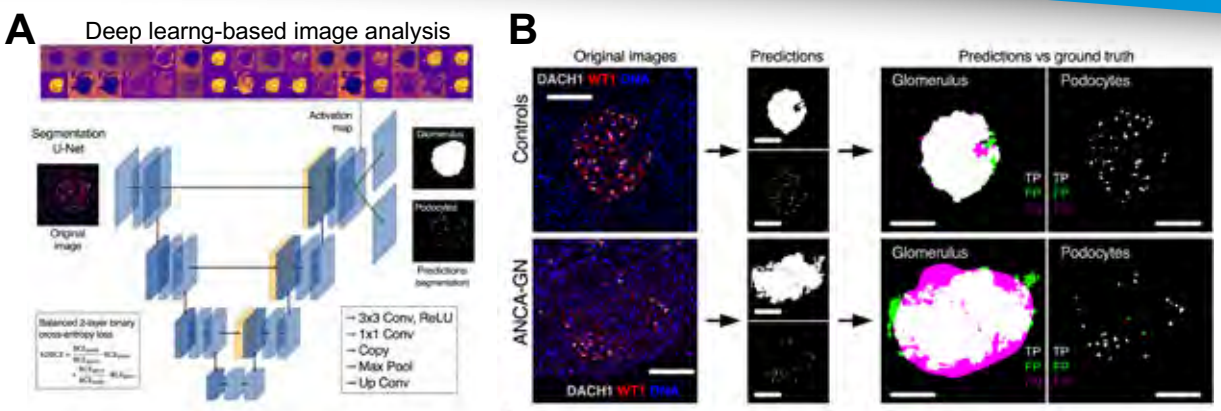


Figure 3: Scheme of the deep learning algorithm and example of images with podocyte morphometrics.
Adapted from Zimmermann *et al.*, 2022.

It is already apparent that the e:Med Fibromap network is making an important contribution to the understanding of fibrosis.

In terms of future prospects, it is particularly valuable that the scientists have found a therapeutic starting point for practically all the areas they are investigating or are already pursuing this in proof-of-principle studies. Fibromap is a good example of a young interdisciplinary team synergistically applying state-of-the-art methods of genomic, imaging and computational models to a cross-disease problem, thereby addressing existing and emerging clinical challenges.

References:

Li, Z., *et al.*, 2023. scMEGA: Single-cell Multiomic Enhancer-based Gene Regulatory Network Inference. Bioinformatics Advances 3-1, 1–5. https://doi.org/10.1093/bioadv/vbad003

Kuppe, C., *et al.*, 2022. Spatial multi-omic map of human myocardial infarction. Nature 608, 766–777. https://doi.org/10.1038/s41586-022-05060-x

Leimkühler, N. B., *et al.*, 2021. Heterogeneous bone-marrow stromal progenitors drive myelofibrosis via a druggable alarmin axis. Cell Stem Cell 28(4): 637–652.e638. https://doi.org/10.1016/j.stem.2020.11.004

Nagai, J.S., *et al.*, 2021. CrossTalkeR: analysis and visualization of ligand–receptor networks. Bioinformatics 37, 4263–4265. https://doi.org/10.1093/bioinformatics/btab370

Jansen, J., *et al.*, 2022. SARS-CoV-2 infects the human kidney and drives fibrosis in kidney organoids. Cell Stem Cell 29, 217-231.e8. https://doi.org/10.1016/j.stem.2021.12.010

Zimmermann, M., *et al.*, 2021. Deep learning-based molecular morphometrics for kidney biopsies. JCI insight 6. https://doi.org/10.1172/jci.insight.144779

Publications e:Med Fibromap: https://www.sys-med.de/de/juniorverbuende/fibromap/publikationen/

Contact:

Prof. Dr. med. Rafael Kramann
Dept. of Nephrology and
Clinical Immunology
RWTH Aachen
kramann@ukaachen.de
www.kramannlab.com

Prof. Dr. Dr. Rebekka Schneider
RWTH Aachen
Div. of Cell Biology
reschneider@ukaachen.de
www.schneiderlab.org

Prof. Dr. Ivan G. Costa
Institute for Computational Genomics
RWTH Aachen
ivan.costa@rwth-aachen.de
www.costalab.org

Prof. Dr. med. Victor Puelles
Center for Internal Medicine
III Medical Clinic and Polyclinic
University Medical Center
Hamburg-Eppendorf
v.puelles@uke.de
https://hckh.org/teams/pis2/victor-puelles

Photos: Rafael Kramann, private; Rebekka Schneider, RWTH Aachen Press/Photo Office; Ivan Costa, RWTH Aachen Press/Photo Office; Victor Puelles, private

a strong hospital network is growing

How HiGHmed Association is advancing the digital exchange of health data

by Eva König and Anna-Maria Zierenberg

How it all started.

HiGHmed is a large scale national medical informatics project within the Medical Informatics Initiative (MII). Its goal is to improve the digital accessibility of medical patient data for clinical research, teaching and health care by establishing Medical Data Integration Centers (MeDICs).

Quickly it became clear that the solutions developed in the network and the bundled expertise of the HiGHmed partners should be used beyond the end of the federal funding. Thus, in 2019, HiGHmed Association was founded as a registered association.

From HiGHmed for HiGHmed.

Since the *MII concept phase* (2016 – 2017) the HiGHmed Consortium has grown. Starting out with its initial three university medical sites HiGHmed counts 14 clinical sites to this date going beyond boarders with its international partners. Local MeDICs are being established at all HiGHmed sites. This gain in experience has enabled HiGHmed Association to specialize in efficiently integrating new partners into the processes of the HiGHmed network. The association not only takes care of the administrative onboarding of the new site, including the introduction to the governance as well as information and document structures. It also provides technical advice on setting up a MeDIC. In addition, HiGHmed Association heads the HiGHmed *Rollout* Working Group, which introduces the various (MII) working groups and use cases to the new partners.

Next to these services – which are focused on MeDIC development – HiGHmed Association supports its members as a central service provider in other matters, for example in the publication of joint tenders or in public relations work. With the organization of symposia, workshops or seminars, HiGHmed Association on the one hand promotes the transfer of knowledge within the network. On the other hand HiGHmed Association also paves the way for new collaborations with partners from industry, science or politics as well as citizen and patient-oriented organizations to further disseminate and consolidate HiGHmed results outside its network.

The MII was launched in 2015 by the German Federal Ministry of Education and Research (BMBF). The funding program aims to advance digitization in medicine by facilitating the exchange and sharing of data between research institutions and healthcare providers through innovative IT solutions.

A MeDIC is an organizational unit of a facility/institution that provides patient data and, if applicable, biomaterials (in collaboration with a biobank) for the purpose of medical research or performs analysis methods and routines.

Team of the HiGHmed Association
Photo: © HiGHmed Association

Overview of projects with participation of HiGHmed Association (Abstract)

In **NUM-RDP** (Routine Data Platform) – a platform for merging COVID-19 research data – HiGHmed Association is mainly responsible for the operation of the central routine data platform. All 34 university medical sites in Germany are successfully connected and can transfer research data. NUM-RDP will be extended beyond its original focus on COVID-19 data with additional funding beginning in 2023.

TRANSIT (daTa pRovisioning And iNformaticS InitiaTive) develops a data management unit for the MII. This unit serves as the technical core for an overarching coordination and provision of datasets from the federated MeDICs to the centrally requesting researchers. HiGHmed Association contributes the expertise gathered in the NUM project and thus avoids redundancies in the design of the national MII/NUM research data infrastructure. In addition, HiGHmed Association supports the sustainability of the project by enabling long-term economically sustainable operation through the development of operating and business models.

HEALTH-X dataLOFT in GAIA-X focuses on building a citizen-centric health data space as a Legitimized, Open and Fundable dataLOFT platform according to GAIA-X standards. Citizens are focused on and should have access to and control over their personal health data, regardless of its origin. HiGHmed Association supports the development of operator and business models for the subsequent, permanent operation of the platform.

The Digital Hub **CAEHR** (CArdiovascular diseases – Enhancing Healthcare through cross-sectoral Routine data integration) focuses on improving the care of cardiovascular patients by standardizing and structuring health data collected from outpatients and inpatients. A research-compatible electronic record should enable communication and the timely delivery of relevant health information for all stakeholders along the treatment process. HiGHmed Associaton manages the external communication activities for the CAEHR project, utilizing the expertise of the HiGHmed Consortium. The association is also responsible for the coordination, implementation and operation of exploitation and dissemination activities to ensure that the project developments will be sustained after the end of the funding.

Partner since:
2020 (founding member)
2020
2022

BERLIN
Charité – Universitätsmedizin Berlin

COTTBUS
Carl-Thiem-Hospital Cottbus

GÖTTINGEN
University Medical Center Göttingen

HANNOVER
Hannover Medical School

HEIDELBERG
University Hospital Heidelberg

KIEL
University Hospital Schleswig-Holstein

KÖLN
University Hospital Cologne

LÜBECK
University Hospital Schleswig-Holstein

MÜNSTER
Westphalian Wilhelms University of Münster

WÜRZBURG
University Hospital Würzburg

BIELEFELD
University of Bielefeld

OLDENBURG
University of Oldenburg

LUXEMBURG
Luxembourg Institute of Health

STUTTGART
Robert-Bosch-Hospital Stuttgart

Location map with all HiGHmed Association members: eleven university medical sites and the Carl-Thiem-Hospital Cottbus, which is being expanded into a university medicine; the Robert-Bosch Hospital in Stuttgart as the first non-university site to become a member of the association; the Luxembourg Institute of Health completes this strong network as the first international partner. Further discussions on membership are being held with interested institutions from science and industry, as well as representatives of patients (Graphic: HiGHmed Association).

What do we do?

- We think beyond the funding periods in order to perpetuate the structures and solutions developed in HiGHmed.
- We develop concepts to expand the existing research areas using models, methods and technologies developed in HiGHmed.
- We facilitate the entry for new collaborative partners on an organizational, legal as well as technical level.
- We enter into strategic collaborations with institutions from science, politics and industry as well as citizen- and patient-oriented organizations; the aim is to expand the HiGHmed network beyond university medicine.
- We provide central services for the HiGHmed network, for example in the areas of public relations, tender offers and knowledge transfer.
- We support the management, coordination and implementation of further research initiatives and projects within the HiGHmed ecosystem.
- We stand by our members in an advisory capacity, both for operational and strategic issues.

From HiGHmed for the Medical Informatics Initiative.

Beyond the HiGHmed Consortium, HiGHmed Association contributes its project competence and experience to various funding projects. One example is the nationwide Network University Medicine (NUM). In the project NUM-CODEX, a secure, expandable and interoperable platform for the central storage and provision of COVID-19 research data was developed using HiGHmed technologies. In the immediate follow-up project NUM Routine Data Platform (NUM-RDP), HiGHmed Association is responsible for the further development of this central platform in addition to its operation. In close cooperation with the already existing MeDIC infrastructure of the MII, all 34 university medical sites in Germany were successfully connected in a short time. This active contribution of HiGHmed to the joint NUM/MII infrastructure will allow generating strong synergy effects. Redundant concept designs and developments should be avoided.

Within the TRANSIT project, HiGHmed Association is pursuing the goal of expanding the MII infrastructure during the current *Consolidation and Extension phase* (2023-2026) by developing and integrating a so-called Data Management Unit. As a component of the federated MII infrastructure, the Data Management Unit will support scientists in their role as data consumers in the future. It offers functionalities for project-specific and temporary consolidation of federated data. Close cooperation with

"I am proud that we founded HiGHmed Association in 2019: It has been growing continuously in terms of projects and employees ever since, and has been able to establish itself as a point of contact for all central questions regarding HiGHmed. This is unique within the Medical Informatics Initiative. HiGHmed Association thus makes a real difference when it comes to win new partners and open up new fields of application, implementing our innovative solutions sustainably and developing them further in the long term."

Prof. Dr. Roland Eils
Chairman of the Management Board HiGHmed Association

Photo: David Ausserhofer / BIH

the German Portal for Medical Research Data (FDPG) is an immanent requirement for this. The Association is thus committed to a uniform appearance of the MII as well as to an optimized user participation and experience.

HiGHmed Association was founded as an institution to perpetuate funding project results. It also advocates sustainable utilization concepts for solutions and techniques that have already been successfully implemented. It is therefore involved in other projects, such as the Digital Hub **CAEHR** or **HEALTH-X dataLOFT** (overview page 81). HiGHmed Association takes on the mission of defining and evaluating alternative business models and operator platforms for building and providing medical platforms based on the expertise from the HiGHmed Consortium.

Our vision.

All these challenges are best met together with progressive concepts and courageous inventiveness. As HiGHmed Association, we support our members, the University Medicine Network and the Medical Informatics Initiative in developing innovative solutions to link patient care and research even better in the future – for a future-proof, data-based and patient-centered healthcare system. We want to make a strong contribution to establishing new technologies in healthcare in a timely and sustainable manner – for the benefit of patients and to support medical and nursing staff.

Profile

Name: HiGHmed Association
Headquarters: Heidelberg
Founded: 2019
Legal form: registered association (e.V.)
Board of Directors: Prof. Dr. Roland Eils (Chairman), Prof. Dr. Dr.-Ing. Michael Marschollek (Deputy Chairman), Prof. Dr. Björn Bergh, Dr.-Ing. Steffen Ortmann, Prof. Dr. Ramin Yahyapour
Director: Stefan Becker
Members: 13 clinical institutions (as of March 2023)
Staff: 6 (as of March 2023)
Goal: Sustainable use of HiGHmed project results for a data-based and patient-centered healthcare system

How to reach us.

Do you have questions or are you interested in becoming part of our network?

Then feel free to contact us:
HiGHmed e.V.
Langer Anger 7-9
69115 Heidelberg
geschaeftsstelle@highmed.org

www.highmed.org

more than an artificial network

The modelling network for severe infectious diseases (MONID) introduces itself

by Jan Ole Berndt, Veronika Bierbaum, Tim Conrad, Jan Hasenauer, André Karch, Mirjam Kretzschmar, Martin Kühn, Alexander Kuhlmann, Berit Lange, Neele Leithäuser, Rafael Mikolajczyk, Kai Nagel, Viola Priesemann, Markus Scholz and André Calero Valdez

Modelling proved to be an important tool for pandemic control during the SARS-CoV-2 pandemic, providing early and relevant insights into transmission dynamics and the effects of behavioral changes or containment measures. In May 2022, the Modelling Network for Severe Infectious Diseases (MONID) was established. The objectives of MONID are the further development of methods in this field and their application for quantitative research into the spread and containment of infectious diseases, the strengthening of interdisciplinary exchange between the modelling groups involved in the network, surrounding disciplines and decision-makers, and the promotion of young scientists.

Table 1a:
Overview of MONID's research priorities

infoXpand

The COVID 19 pandemic was accompanied by an "infodemic", i.e. an excess of information about the virus, protective measures and government guidance. In particular, the mis- and disinformation and conspiracy theories spread on the internet were blamed for the increasing polarisation of opinion, radicalisation and declining trust in institutions. The main objective is to understand the feedback loop between (mis)information spread and pandemics and to derive strategies for future crises.

INSIDe

Development of a simulation platform that incorporates both test-based infection data and wastewater monitoring to provide an accurate picture of the infection situation. For infection dynamics and sewer networks, hybrid models are used that allow more detailed observations at the point of interest through multi-scale modelling. At the same time, the platform created should be able to monitor several pathogens simultaneously.

MODUS

Operation and further development of an agent-based simulation model of virus spread based on mobile phone data, in particular for predicting the effects of measures. Development of mathematical methods for model reduction and efficient model simulationto achieve a significant reduction in computing time.

OptimAgent

Development of an agent-based simulation platform to support decision-making regarding the prevention and control of infectious diseases, taking into account heterogeneity in intra- and inter-individual contact patterns, mobility, individual sociological and psychological characteristics and socio-economic factors.

Figure 1: Summer School 2022 (Photo: University Medicine Halle (Saale)).

Aims and structure of the modeling network

The Modeling Network for Severe Infectious Diseases (MONID; https://webszh.uk-halle.de/monid/?page_id=787&lang=en), funded by the German Federal Ministry of Education and Research, is an interdisciplinary consortium of more than 30 university and non-university institutions in Germany conducting research on population-based modeling of infectious disease transmission and associated health consequences in seven consortia. The first cross-network activities (see below) already show the substantial added value created by the networking of the modeling groups and their research activities to strengthen modeling competence in Germany.

The modeling consortia involved in MONID bring together experts from a wide range of disciplines, which is reflected in the complexity of their research focus and goals (see Table 1a and 1b for details) and in the broad spectrum of modeling methods applied. In addition, the consortia apply diverse existing data sets (e.g., contact surveys, GPS mobile phone data, data from population-based cohorts, social media and wastewater moni-

Table 1b: Overview of MONID's research priorities

PROGNOSIS

Develop models to predict hospital burden at multiple levels of care during an epidemic, assess the impact on the health system, supply chains and human resources. Building a freely available and extensible modelling approach to provide predictions of hospital burden and assess the effectiveness of counterstrategies.

RESPINOW

The overall goal is to support the control of pandemic and endemic respiratory infections and to better understand the dynamics of these infections and their interaction with each other and with non-pharmaceutical interventions. To this end, evidence syntheses will be conducted, data collected in population-based cohorts on the dynamics of RSV, influenza and pneumococci during and after the pandemic, integrated modelling of medium- and long-term effects of non-pharmaceutical interventions will be established, and platforms with short-term predictive models of these pathogens will be built.

SEMSAI

The aim of SEMSAI is to explore the impact of communicating predictions on perceived risks and future behaviour of the population and to incorporate this into modelling. The consortium consists of three partners from the fields of psychology and social science disaster research, mathematical modelling and prediction, and agent-based social science.

toring, health insurance data) and will additionally collect new data, e.g., in experiments or panel surveys, to adequately inform the models. Important strategic goals of MONID in this context are the harmonization of methods between the modeling networks as well as the development of common platforms for internal communication, data exchange and management in order to make optimal use of synergy potentials.

A coordination office coordinates and implements the cross-network activities (e.g. joint workshops, conferences, external communication/public relations, data sharing). In addition, a steering group was established as a joint management structure of the collaborations, which is responsible for the overall strategic orientation and methodological improvements. Through the coordination office, the steering group is the central contact for the modeling network for scientific inquiries. The structures created aim to harmonize communication with decision-makers and the public and to actively promote networking with surrounding disciplines and research initiatives. The respective research foci and goals of the modeling networks are summarized in Table 1.

Interdisciplinary activities of the modeling network

Summer School

The joint MONID Summer Schools are a central activity for the promotion of young scientists. The first Summer School, which took place in Halle (Saale) from September 12-16, 2022, focused in particular on the fundamentals of mathematical modeling of infectious disease spread and strengthened the networking of the seventeen participating PhD students from different disciplines (e.g., mathematics, physics, epidemiology, psychology, economics).

Topics included epidemiological principles, model types, optimization/calibration, implementation of interventions, health economic evaluations, and visualization and communication of model results. The Summer School 2022 was divided into units with lectures and practical hands-on sessions in which program codes were developed. National and international experts were invited for the presentations, e.g. from the London School of Hygiene & Tropical Medicine, University of Oxford, Hasselt University.

In August 2023 a Summer School on the interaction of pandemics and infodemics took place in Lübeck.

Structured scientific exchange on contact behavior and social networks

Individual contact behavior and the corresponding social network structures are major factors influencing the transmission of infectious diseases and are therefore a core element of modeling efforts. During the SARS-CoV-2 pandemic, new data on the dynamics of individual contact structures and behavioral changes were collected. In addition, important insights were gained into the role of aerosols in respiratory disease transmission. GPS mobile phone data came into focus as an additional data source for assessing contact intensities and networks.

Due to its central importance for modeling, a structured scientific exchange on data sources, modeling of contact behavior and the importance of social networks was established within MONID to gain new insights in this field and to further develop methods for implementing social network structures in models of infectious disease spread. To this end, five working groups have been formed to address the following topics: a) integration of social contacts and aerosol dispersal to model transmission processes, b) integration of mobility data, contact rates and durations to model transmission processes, c) influence of cluster structures in contacts on dynamics of infectious disease spread, d) methods to adjust contact networks with respect to specific pathogen characteristics, e) methods to incorporate contact information from agent-based models into differential equation models.

Figure 2: Summer School 2022 (Photo: University Medicine Halle (Saale)).

Workshops on modeling SARS-CoV-2 scenarios for the winter of 2022/2023

Three possible scenarios for the further course of the SARS-CoV-2 pandemic in Germany in the winter of 2022/2023 were simulated during two workshops. In total, seven modeling groups of the network participated. Objectives of the workshops were the comparison of modeling techniques, parameterization and results, since the models developed are based on different approaches and represent the infection event in varying detail. The diversity of applied methods should enable a more robust evaluation of the scenarios. Current estimates for protection levels against SARS-CoV-2 infection and severe COVID-19 from the IMMUNBRIDGE study were included in the modeling efforts (Lange *et al.,* 2022).

- **Scenario 1:** Omicron variants BA.4/BA.5, which caused the 2022 summer wave in Germany, will continue to dominate in the winter of 2022/2023.
- **Scenario 2:** A new variant that can partially escape previously acquired immune protection in the population, thereby increasing the risk of infection (but not the risk of severe disease progression), will become prevalent in the winter of 2022/2023.
- **Scenario 3:** As in scenario 2, a new variant will prevail, but with an increased risk of severe disease progression: the probability of hospitalization is 2-3 times higher compared to the omicron variants BA.4/BA.5.

For scenarios 1 and 2, most of the simulations showed a wave of infection during the winter that would most likely not lead to an excessive burden on hospitals, i.e. the peak values of the BA.1/2 wave in winter 2021/2022 would not be exceeded. In scenario 3, the peak hospital burdens of SARS-CoV-2 reached so far would probably be significantly exceeded even with intensified vaccination campaigns. However, this scenario was viewed as rather unlikely. More detailed results can be found in the corresponding statement of the modelling network (Berndt *et al.,* 2022).

Despite very different modelling approaches and variations in parameterization - no restrictive specifications were deliberately made for model parameters in order to achieve a certain variability in the results - the simulations showed a comparatively uniform picture. If there were substantial deviations in the results, these could be explained by the parameterization. For example, simulations in which a very strong seasonality was assumed in connection with a relatively high immunity waning also led to significantly higher hospital loads in scenarios 1 and 2.

Overall, the combination of the different modelling approaches proved to be very valuable, as it enabled an external validation of the individual models and thus more robust assessments of the epidemiological situation of SARS-CoV-2 in Germany against the background of the complex and heterogeneous immunity situation in Germany.

Regular meetings on the epidemiological situation of severe respiratory diseases

Building on the two workshops on modelling SARS-CoV-2 scenarios for autumn/winter 2022/2023, a bi-weekly online meeting was established to discuss the current epidemiological situation, to plan, coordinate and review joint modelling activities of the network, and to draft joint statements.

During the meetings, new SARS-CoV-2 scenarios were simulated for the winter of 2022/2023, involving the lifting of the isolation requirement and the mask requirement in public transport. The results of the modelling suggested that lifting the isolation and masking requirements would moderately increase the disease and SARS-CoV-2 burden on the health system compared to maintaining these measures. In this second MONID statement (Bierbaum *et al.,* 2022), the impact of lifting the isolation and masking requirements in public transport on the spread of other severe respiratory diseases in November/December 2022 were also discussed, in particular the exceptionally high hospital occupancy rates among children combined with a strained staffing situation in pediatric hospitals.

References:

Berndt, J.O., Conrad, T., Hasenauer, J., Karch, A., Kheifetz, Y., Kirsten, H., Krueger, T., Kühn, M.J., Kuhlmann, A., and Lange, B., *et al.* (2022). Szenarien für den Verlauf der SARS-CoV-2-Pandemie im Winter 2022/23 - Ergebnisse eines Workshops des Modellierungsnetzes für schwere Infektionskrankheiten (Modellierungsnetz) (Zenodo https://zenodo.org/record/7126032#.ZD5R9LrP1aQ).

Bierbaum, V., Calero Valdez, A., Conrad, T., Hasenauer, J., Karch, A., Kuhlmann, A., Krueger, T., Lange, B., Loeffler, M., and Mäs, M., *et al.* (2022). 2. Stellungnahme des Modellierungsnetzes für schwere Infektionskrankheiten (MONID) (Zenodo https://zenodo.org/record/7440730#.ZD5RqLrP1aQ).

Lange, B., Jäger, V.K., Rücker, V., Harries, M., Hassenstein, M.J., Dreier, M., Holt, I. von, Budde, A., Kurosinski, M.-A., and Armann, J., *et al.* (2022). 2. Interimsanalyse des IMMUNEBRIDGE-Projektes zur Kommunikation von vorläufigen Ergebnissen an das Modellierungsnetz für schwere Infektionskrankheiten (Zenodo https://zenodo.org/record/7177592#.ZD5RZbrP1aQ).

Contact:

Ass.-Prof. Dr. Alexander Kuhlmann
Head of the Coordination Office of the Modeling Network for Severe Infectious Diseases
Martin-Luther University
Halle-Wittenberg
Medical Faculty
Halle (Saale)
Alexander.Kuhlmann@uk-halle.de

Dr. Berit Lange
Speaker of the Steering Group of the Modeling Network for Severe Infectious Diseases
Helmholtz Centre for Infection Research
Braunschweig (HZI)
Department of Epidemiology
Braunschweig
Berit.Lange@helmholtz-hzi.de

Prof. Dr. Rafael Mikolajczyk
Speaker of the Steering Group of the Modeling Network for Severe Infectious Diseases
Martin-Luther University
Halle-Wittenberg
Institute for Medical Epidemiology, Biometry and Informatics
Halle (Saale)
Rafael.Mikolajczyk@uk-halle.de

https://webszh.uk-halle.de/monid/?page_id=787&lang=en

Photos: Alexander Kuhlmann, University Medicine Halle (Saale); Berit Lange, HZI Verena Meier; Rafael Mikolajczyk, University Medicine Halle (Saale)

the de.NBI network

Continuation of activities as an established bioinformatics infrastructure

by Alfred Pühler, Alexander Sczyrba and Andreas Tauch

The German Network for Bioinformatics Infrastructure (de.NBI) was established between 2015 and 2021 as part of a funding measure of the Federal Ministry of Education and Research (BMBF) with funding amounting to almost 100 million euros. After a seven-year project duration the good news came in December 2020 that the Committee on Economic Affairs of the German Bundestag decided to carry out the continuation of the German Network for Bioinformatics Infrastructure in the Helmholtz Association at Forschungszentrum Jülich. This article deals with the impact of this decision on the activities of the de.NBI network and with the associated structures of the German ELIXIR node and the computing infrastructure de.NBI Cloud.

The de.NBI network – an infrastructure for the promotion of bioinformatics activities in Germany

The central task of de.NBI has always been to support experimental researchers in the life sciences in the analysis of large amounts of data. This is done by providing proven bioinformatics tools and extensive training opportunities for in-depth familiarization with the analysis programs. However, "cloud computing", "international collaboration" and "industry connection" (Figure 1) also represent important areas of work. See the article in systembiologie.de (04/16).

The annual federal funding of €10 million available since 2022 ensures that de.NBI's areas of activity will continue. The state of NRW is additionally supporting de.NBI with € 1.1 million annually and the other federal states with de.NBI sites have also been asked to contribute financially to de.NBI. A consortium agreement is currently being drawn up to regulate cooperation with the de.NBI partners, the participating federal states and Forschungszentrum Jülich as the coordinating body. Another sign of life was given by the de.NBI members by founding the de.NBI association. The de.NBI e.V. aims to continue the tasks of the de.NBI network in order to maintain the achievements of the previous BMBF funding measure. On the working level, however, the established de.NBI network has continued its activities. Thus, as usual, a large number of training courses are offered and also the series of de.NBI summer/winter schools is continued with current topics. The outstanding activities of de.NBI, however, are the scientific design of the German ELIXIR node and the operation of the federated de.NBI Cloud.

Figure 1: Key data for the de.NBI network, the de.NBI Cloud and the German ELIXIR node (Source: de.NBI office.).

The de.NBI network – starting point for the successful establishment of the German ELIXIR node

Since 2016, Germany has been a member of the European bioinformatics infrastructure ELIXIR, an intergovernmental organization that brings together life science resources such as databases, software tools, training and educational materials, and cloud storage from across Europe. The goal of ELIXIR is to coordinate these resources across member states to form a single European bioinformatics infrastructure. This should make it easier for scientists to efficiently use the huge amounts of data currently being produced in the life sciences to conduct data-driven research.

As a national ELIXIR node, ELIXIR Germany is run by de.NBI members of 21 contractual partners from German universities and research institutions (Figure 1). The change in funding has only a minor impact on ELIXIR Germany, which still offers more than 100 bioinformatics tools for the analysis of life science data. In the future, researchers will be trained in data analysis in accompanying training events. As usual, ELIXIR Germany scientists participate in internal ELIXIR research projects to harmonize the national service offering with the European partners.

The de.NBI network – home of the de.NBI Cloud

The de.NBI Cloud, which has been successively built up since 2016, will also be available to scientists in the usual scope and quality after the continuation. The use of the de.NBI Cloud is free of charge for scientists from Germany, ongoing projects will continue to be supported, and new ones will be approved. The need for and success of the de.NBI Cloud is reflected in the steadily increasing number of users: of the more than 2,200 users, over 600 registered in 2022 alone, who are now active in over 730 projects. With over 56,000 CPUs and more than 100 petabytes of storage, followed by more than 500 publications the de.NBI cloud is an essential resource for the data-driven life sciences. This is true for many National Research Data Infrastructures (NFDI) projects, such as NFDI4Microbiota, GHGA, NFDI4Biodiversity, or DataPLANT, working on limited computing infrastructure and relying

on the de.NBI Cloud. Nevertheless, the long-term funding of most de.NBI Cloud sites is not yet guaranteed and represents an open problem.

Further information: www.denbi.de

Contact:

Prof. Dr. Alfred Pühler
Founding Coordinator of the
German Network for Bioinformatics
Infrastructure (de.NBI)
Center for Biotechnology
University of Bielefeld
puehler@cebitec.uni-bielefeld.de

Prof. Dr. Andreas Tauch
Head of Node ELIXIR Germany
Center for Biotechnology
University of Bielefeld
tauch@cebitec.uni-bielefeld.de

Prof. Dr. Alexander Sczyrba
Speaker of the de.NBI Cloud
Center for Biotechnology
University of Bielefeld
sczyrba@cebitec.uni-bielefeld.de

INTERVIEW WITH OLIVER STEGLE, CO-FOUNDER AND SPEAKER OF GHGA

GHGA: the german human genome-phenome archive

A national omics data infrastructure in an international context

by Ulrike Träger, Jan Eufinger and Nicole Schatlowski on behalf of the GHGA-Consortium

Since the end of 2020, an interdisciplinary team has been working to establish the German Human Genome-Phenome Archive (GHGA) as part of the National Research Data Infrastructure (NFDI). GHGA bundles the expertise from more than 20 universities, Helmholtz centres and other research institutions to develop an infrastructure for sharing human genome data in accordance with the "FAIR principles".

In this interview, GHGA co-founder and speaker Oliver Stegle explains why sharing human omics data is such an important undertaking. Besides his contributions to GHGA, Stegle and his team at the German Cancer Research Center (DKFZ) and the European Molecular Biology Laboratory (EMBL), develop new statistical and machine learning methods to decipher molecular variation across individuals, space and time.

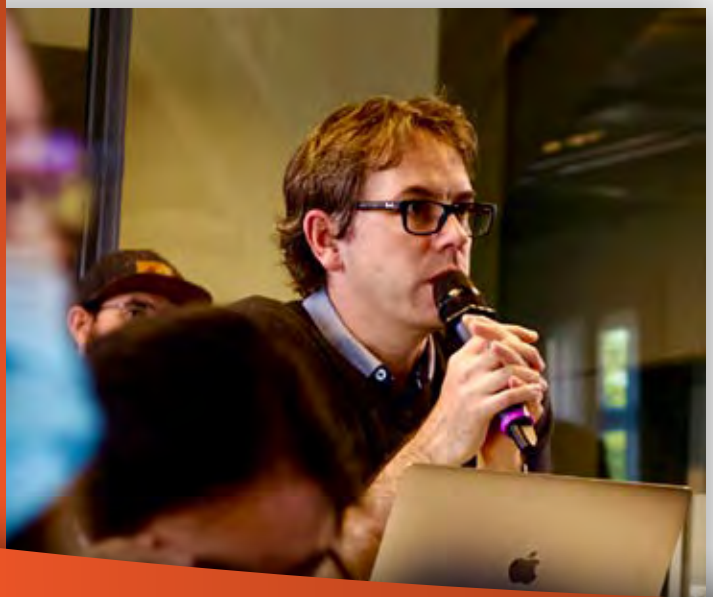


Figure 1: Oliver Stegle in discussion at the GHGA Annual Meeting 2022 in Tübingen (Photo: © Ulrike Träger).

gesundhyte.de: *Thank you for joining us today, Herr Stegle. Let us start with your motivation, what drives your research?*

Prof. Dr. Oliver Stegle: Thank you for having me. My passion is to solve scientific puzzles. In my lab, we develop computational approaches to understand genomic variation and decipher their consequences. Our methods contribute to uncover genetic variations that influence our risk for developing certain diseases or are generally associated with human traits.

This scientific endeavour is closely linked to the possibility to actually use genome data and have access to population cohorts – including studies on cancer and other diseases. Therefore, for some years now, in parallel to our research activity, I have been working to make genome data more usable for research.

gesundhyte.de: *Is this commitment also the birthplace of GHGA?*

Prof. Dr. Oliver Stegle: Yes. A really fantastic community of researchers and clinicians in Germany have come together to create the German Human Genome-Phenome Archive. An effort to create a systematic and well-structured infrastructure for genomic data to enable exciting research projects with human omics data.

Data is the fuel for science and AI research. We need datasets to train complex algorithms based on artificial intelligence, but also to validate and test new methods so we can say:

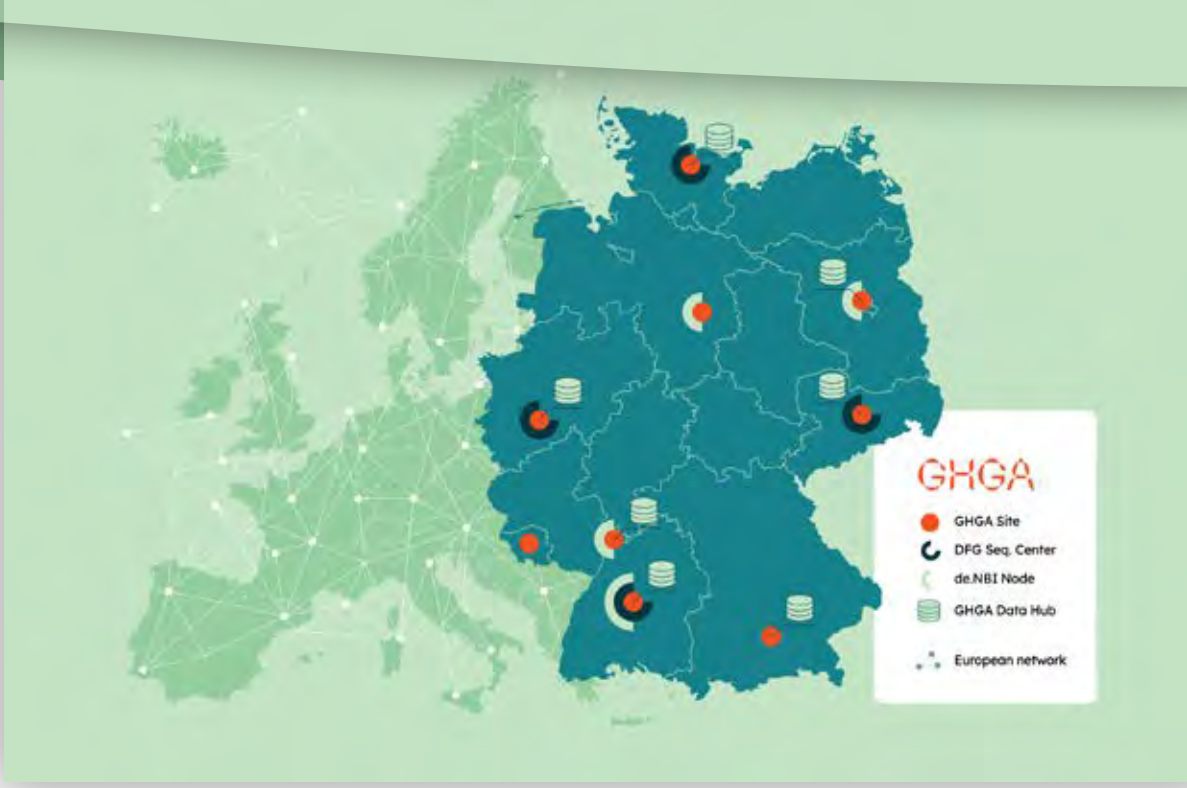


Figure 2: GHGA consists of interconnected data hubs - linked to leading genomic medicine institutions, large omics data producers, high-performance computing centres and existing cloud infrastructures (e.g. within de.NBI / ELIXIR-DE). GHGA is part of the federated European Genome Archive (fEGA) and involved in the 1+Million Genomes Initiative (graphic design: DIE KAVALLERIE GmbH, copyright: DKFZ/GHGA).

"Yes, we really understand a piece of the genome puzzle and how it affects health".

In other countries, there are already impressive resources, such as the UK Biobank. It provides data from half a million British people for research. Genomic and phenotypic data: everything from DNA sequences, health outcomes, dietary preferences to their COVID status. And this is where it gets exciting, with the prospect of replicating similar ideas and structures in Germany.

gesundhyte.de: *So if we go back in time: How did the idea to build a data infrastructure in Germany emerge?*

Prof. Dr. Oliver Stegle: It was quite a funny story. We were a group of colleagues who all left Cambridge at the same time moving to Heidelberg. My colleague next door was setting up a mouse facility – creating a cohort of mice at different ages.

„A really fantastic community of researchers and clinicians in Germany have come together to create the German Human Genome-Phenome Archive."

It was a great resource, which he wanted to share with other local research groups. I felt like I had to do something for the community in omics research as well. An effort to do more than just science, but also to open up new opportunities in the long term.

The concrete idea then developed over dinner with Oliver Kohlbacher, with whom I co-founded GHGA. We felt that we needed to bring these ideas that are already established elsewhere, e.g. in the UK and Spain, where the European Genome-Phenome Archive (EGA) has been built and is being run, to Germany. Our vision was to make them usable for our country and to improve them even further together with many partners.

gesundhyte.de: *Making the GHGA infrastructure work at a national level - what is special about Germany?*

Prof. Dr. Oliver Stegle: Many European countries face similar problems, but each has its own story. In Germany, one of the real opportunities is that many researchers have been leaders in the field of genome research for the last twenty years. There's been fantastic pioneering work in oncology, for example, and German researchers participate in major international efforts – both from an analysis perspective and the availability of the generated data. This is an excellent starting point and a track record to build on.

The challenges in Germany, as in most countries, are to an extent cultural: Historically, the mindset to share data is perhaps stronger in some Anglo-Saxon countries than in Germany.

The other major challenge is – not surprisingly – the, in comparison to other countries, restrictive interpretation of the EU-wide General Data Protection Regulation (GDPR) in Germany. This is further complicated by the federated structure and the associated state-specific implementation of data protection laws.

It's not that we can't do what others do. It is just often more complicated to justify and achieve the appropriate legal basis for a particular data processing task because of the diversity of systems and the associated processes and paperwork.

gesundhyte.de: *Work on GHGA started at the end of 2020, where does it currently stand?*

Prof. Dr. Oliver Stegle: The project has been running for more than two years now, and it's important to realise that building these infrastructures is a marathon rather than a sprint. We have made exciting progress in a number of areas, but it takes time.

One dimension, that is key to many users, is the archive function. The ability to deposit data in a secure way in order to be able to share it. We have just launched the infrastructure's first service to help make data from Germany discoverable: the GHGA Metadata Catalog (http://catalog.ghga.de). By the end of 2023, we plan to launch the first fully functional instance of the GHGA archive functionality and offer researchers at national level a secure way to deposit and share data.

Importantly, GHGA is neutral when it comes to managing access to submitted data. Submitters can come from all over the country. We can accommodate a wide range of different data governance models that have emerged through different studies or research projects. This is a very important practical detail and will hopefully help to boost acceptance.

gesundhyte.de: *In addition to software developers, GHGA has a large interdisciplinary team. What are some of the complementary efforts GHGA invests in?*

Prof. Dr. Oliver Stegle: Building such an infrastructure is much more than just a piece of software, many other moving parts are involved. We have an ELSI (*ethical, legal and social implications*) working group that has developed standards and procedures on how to think about consent and how to structure it. This gives us the legal basis to deal with datasets deposited in archives (such as GHGA or other related activities) and making them available for secondary research.

We are also thinking extensively about analysis workflows and standardisation. GHGA is not just about making data available, but also about structured and standardised approaches for its analysis. It's about bringing the community together and creating standards for the future.

GHGA Resources:

- Data infrastructure for GDPR-compliant sharing of human omics data for secondary research use
- Standardised, interoperable, and reproducible omics workflows for the research community
- Legal and ethical foundations for genomics research
- Metadata model to provide standardised information on submitted genomic data and facilitate data findability
- Training material for and about genomics research and its relevance for society

for more information see https://www.ghga.de

The german-language GHGA podcast sheds light on questions about the Code of Life: DNA and genome research.

gesundhyte.de: *Finally, let's venture into the future: where do you see GHGA in 10 years?*

Prof. Dr. Oliver Stegle: The exciting question is how we can link and integrate results from research with clinical care even better.

How can we create an even more active exchange? How can we translate more ideas from research into health care?

One example is the use of genomic data as a basis for therapeutic decisions, e.g. in molecular tumour boards. Sequencing data can provide vital evidence for the selection of personalised treatment strategies. In turn, the data can be further used for research – with the patient's consent. The integration of these datasets will in turn help to develop new treatment options that will benefit future patients. This approach will also be a central part of the soon-to-be-launched national genome initiative genomDE, in which GHGA is also involved.

We are convinced that GHGA can make an important contribution and be a piece in the puzzle to tie these two worlds, patient treatment and research, even closer together. We are already looking forward to being able to include the first data from exciting studies in the near future and thus make them available for research on new therapies and diagnostics.

References:

1. https://www.ghga.de
2. http://catalog.ghga.de
3. Mark D Wilkinson, Michel Dumontier, I Jsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, *et al.*: The FAIR Guiding Principles for scientific data management and stewardship. 1996, doi:10.1038/sdata.2016.18 https://www.nature.com/articles/sdata201618

Contact:

GHGA Office
Deutsches Krebsforschungszentrum (DKFZ) – W620
Im Neuenheimer Feld 280, 69120 Heidelberg
contact@ghga.de
www.ghga.de

from AI research to AI care

Health and added value through artificial intelligence

by Karsten Hiltawsky and Susanne Boll

Whether in nursing, rehabilitation or outpatient medical care, artificial intelligence (AI) promises enormous potential to improve healthcare. New types of innovation processes and innovative business models are needed to enable patients to benefit from this potential. This involves overcoming technical, ethical, regulatory and economic challenges. Economically viable business models help to bring AI-based innovations to the breadth of the healthcare system and to accelerate the digitization necessary for this.

Experts from the working groups Health Care, Medical Technology, Care as well as Innovation, Business Models and Processes provide an overview of how AI-based business models can be successfully implemented in the healthcare sector in Germany. The focus will be on opportunities, challenges, and the regulatory framework in Germany.

AI as an opportunity for healthcare companies

When analyzing the opportunities of AI applications for healthcare companies, it is important to keep in mind that the medical technology industry in Germany is strongly characterized by small and medium-sized enterprises (SMEs). However, especially for many SMEs and startups, the financing and approval of AI medical devices in healthcare poses major challenges. Specific challenges include the lack of suitable training data, data security issues, lack of AI expertise and confidence in AI technologies, or the unclear prospect of reimbursement. Pre-financing through the hope of increasing market share also appears risky in some cases. This is especially true for comparatively small companies, as the AI economy is a platform economy dominated by large hyperscalers. For the long-term success of AI-based business models, the prospect of new technology reimbursement as well as specialized funding programs would be helpful.

Nevertheless, AI-based business models offer great potential in both the primary and secondary healthcare markets. The secondary healthcare market is becoming increasingly rel-

The input is based on the white paper **KI-Geschäftsmodelle für die Gesundheit – Innovation stärken, Finanzierung gestalten. München**, 2022. The authors are members of the working group Medical Technology, Nursing and of the working group Business Model Innovations. https://doi.org/10.48669/pls_2022-3

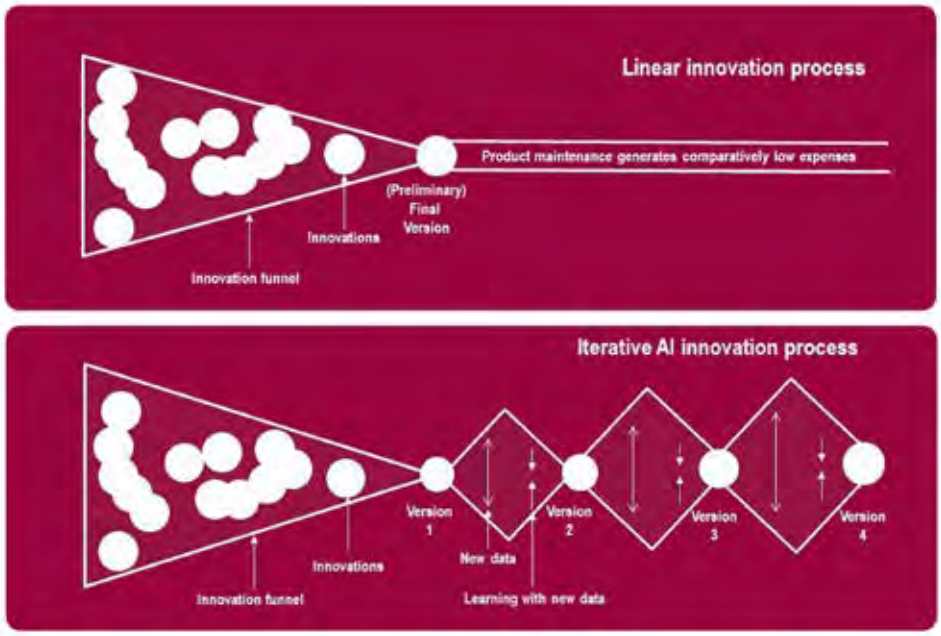


Figure 1: Linear innovation process and iterative AI innovation process (Source: Plattform Lernende Systeme, vgl. Chesbrough, H. W. (2003): Open innovation: The new imperative for creating and profiting from technology. Harvard Business School Press).

evant, which has great potential with meaningful, high-quality AI applications in terms of monetization as well as for personal health and health maintenance. Promising applications for new AI business models for companies in the secondary healthcare market include personal AI-based health applications such as the evaluation of wearables as well as applications in the field of care and rehabilitation. For healthcare companies, AI applications also offer the opportunity to expand their business models by optimizing existing and developing novel offerings with AI.

AI healthcare applications are based on technologically complex AI models and often apply training procedures using numerous clinical data. This cost intensive process must be refinanced via revenue models. The following two revenue models are particularly worthy of mention:

- **Same revenue model for AI-based improvement of an existing offering:**
  There are applications that can improve, facilitate, and automate an existing offering with AI. AI enhances work and process steps but does not fundamentally change them. Typically, companies use such enhancements to differentiate their products. However, the existing business model – such as the sale of a capital good with maintenance and services – remains unchanged.
- **Opportunity of new revenue models for novel AI offerings:**
  The situation is different, for example, in the case of decision support systems (e.g., in the pre-classification of radiological image data): Here, the use of AI enables a novel service. To monetize this, different business models can be examined – away from the licensing business with one-off payments to the „software-as-a-service“ business, rental model, or usage-dependent payments. At first glance, this benefits the providers of such solutions, because they can forecast revenue much more precisely based on their user numbers over a longer period of time. Users also benefit from such service offerings through reduced costs for the maintenance and configuration of the software, for example through the automatic provision and updating of the service via the internet or cloud applications. In addition, the expenses for the services, which replace high one-off financing, are spread over the entire term. Very often, the business models are also supplemented by continuous product maintenance and included services such as protection against cyber-attacks and data loss. In this way, customers can also insist on continuously high quality since they are paying continuously.

AI-specific innovation problems: The causality dilemma

One AI-specific innovation problem is the so-called causality dilemma: The quality of the result depends strongly on the quality and the number of data sets, so that the benefit can often only be proven in the long term. Since the proof of benefit is costly, many companies cannot afford the considerable validation effort. To provide an economic incentive, temporary cost coverage should be discussed until the benefit has been proven. Based on the experience of innovation funding for medical technology and pharmaceuticals, the new German regulation on l**ow-risk Digital Health Applications** (Digitale Gesundheitsanwendungen, DiGA) can serve as a temporary reimbursement option as a model to bridge the time from potential recognition to reliable proof of benefit.

Field of action: Reimbursement and funding

Based on the evaluation of the pathways to reimbursement by the health insurance, it should be discussed whether the existing conditions are already sufficiently tailored to AI-based innovations. Short innovation cycles, the large heterogeneity of applications, and the potential changeability of functionality are just some of the features. It is also unclear whether the existing evaluation processes sufficiently consider the potential of AI. Based on the problem of the causality dilemma, it is apparent that the quality of the technology may increase with more data and training, and thus at the beginning of development the actual benefits may not yet be sufficiently demonstrable. Digital Health Applications (DiGA) already address this issue for low-risk applications with a trial period of one year when the potential benefit is recognized, this could also be extended for high-risk class digital medical devices. The secondary healthcare market also offers a lot of potential. Nevertheless, AI-based examination and treatment methods for which the benefit, medical necessity and economic viability have been proven should be integrated into the primary healthcare market. Only in this way all citizens can benefit from holistic, personalized, patient-centered healthcare. This close link between the primary and secondary healthcare markets requires, among other things, a high level of quality assurance in the secondary healthcare market as well.

Field of action: Data

The availability of data, especially training and validation data for AI, is one of the key prerequisites for successful AI business models in healthcare. Much progress has been made in this field of action in recent years, among other things, treatment data are being made usable for research with the help of coordinated patient consents, and possibilities for intersectoral data exchange are being tested. Within the framework of GAIA-X, a European federated data infrastructure is being created in which health data can be evaluated securely and for distributed AI applications also across countries in the future. These secure and interoperable possibilities for data storage and data use not only create advantages for individual treatment and patient sovereignty. They also enable new applications at the population level, for example in epidemiological questions.

Field of action: Startup funding and secondary healthcare market

Despite some successful examples of German AI health startups, it is apparent that the funding sums generated in Germany are rather low in international comparison. In this context, among other things, the **high investment sums and long waiting times** present themselves as **challenges for SMEs and startups**: In the case of digital business models, a lot of money often must be invested in order to build up and establish platforms. This can be particularly difficult for startups or SMEs. Therefore, from a European and German perspective, ways should be found to bring AI into the mainstream and make it financeable. The main issue here is to create regulatory framework conditions that make it easier to participate in AI health startups with venture capital/corporate venture capital. It would be important, for example, to make it easier for large institutional investors to invest in startups.

Field of action: Innovation and value creation networks

As in other application fields of digital transformation, the aspect of networking is also of central importance for the healthcare sector. A major added value of digitization in the healthcare sector could be the large number of data points from different sources (multiple devices, different employees, different processes in a value chain), which can be related to each other and

thus become the base for a higher level of knowledge. In addition, health status data of a person (holistically or partially) are particularly relevant in the healthcare sector. They can form the base for a digital health twin of a person, which, enriched with information on nutrition, activities, stress, etc., can also enable a simulation of the future state.

Field of action: Certification and liability

When AI is certified as a medical device or diagnostic procedure, it is currently done as standalone software or primarily as part of an overall product. For a medical device to be introduced to the EU market, it must be regulated uniformly throughout the EU, which requires a conformity assessment procedure to prove defined safety, quality, and performance requirements. Depending on the risk class, the respective conformity procedure and the involvement of a notified body differ to independently test and certify the medical device.

The **European Commission's White Paper on AI** points out the danger of unclear regulations on risks, which can lead to legal uncertainty for companies and ultimately reduce their competitiveness. Similarly, users, doctors and patients should not bear the entire risk. For a trained AI system that does not continue to learn during use, there is no significant difference in liability from medical devices.

Liability issues could arise in the more distant future with AI systems that continue to learn. This raises the question of whether the liability system needs to be adapted, at least for AI technologies of high-risk classes. Liability for AI systems in healthcare should therefore clearly consider the special characteristics of AI systems, such as errors in the algorithms, or that AI systems can learn based on trial and error or mistakes.

Field of action: Digital ethics and patient perspective

For these AI technologies to establish themselves on the market, acceptance by patients and users is necessary. Important foundations for creating the trust that is essential for the acceptance of AI in healthcare are the establishment of digital ethics principles and their consideration in the development and application of new digital healthcare solutions, as well as the consideration of the patient perspective along the entire digital value chain. This applies to applications that directly affect patients or are used by them. Trust must be promoted and maintained by all players in the healthcare sector, including private-sector market participants. Only in this way there can be long-term acceptance (and success) of these innovations. Trust arises when the technically feasible is measured against legal and ethical standards. This can be achieved through standards, guidelines or "seals of approval", but above all through proven, transparent action. Players in the healthcare sector should proactively and openly commit to taking these principles into account to promote acceptance of and trust in AI-based innovations. To maintain this trust, AI-based business models must always and first be evaluated against the background of the medical benefit as well as against general ethical and social criteria.

Contact:

Prof. Dr. Susanne Boll
Professor of Media Informatics and Multimedia Systems,
University of Oldenburg
Working Group Leader Innovation, Business Models and Processes,
Plattform Lernende Systeme
susanne.boll@uni-oldenburg.de

Prof. Dr.-Ing. Dr. med. Karsten Hiltawsky
Head of Corporate Technology and Innovation, Drägerwerk AG & Co. KGaA
Working Group Leader Health Care, Medical Technology, Care,
Plattform Lernende Systeme
karsten.hiltawsky@draeger.com

events

NephroCAGE

How artificial intelligence can assist with kidney failure

by Matthieu-P. Schapranow

The German-Canadian consortium NephroCAGE is funded as a lighthouse project by the German Federal Ministry of Economic Affairs and Climate Action since 2021. The aim of the consortium is to demonstrate the added value of artificial intelligence (AI) using a very specific clinical example: kidney transplantation. In the case of insufficient kidney function regular dialysis is necessary: there are currently approx. 100,000 patients in Germany and approx. half as many in Canada on regular dialysis. 1,992 kidney transplantations in Germany (German Organ Procurement Organization) and 1,673 in Canada (Canadian Institute for Health in Canada) were performed in 2021. However, suitable donor organs are rare: in Germany, for example, 6,593 and in Canada 3,060 patients are on a waiting list in the same year. Even after a transplantation there is a risk of complications that can lead to severe limitations in kidney function or, in the worst case, even total loss of the organ.

The NephroCAGE consortium is evaluating the use of AI methods for the analysis of multimodal transplant data. Such AI models should help to detect possible complications after transplantation at an earlier stage and to support nephrologists in their work. For this purpose, the clinical partners from Germany and Canada are providing comprehensive transplant data over a period of up to two decades. These are health data that are particularly sensitive. To this end, the consortium is developing the NephroCAGE Federated Learning Infrastructure (FLI) specially tailored to the requirements. Thus, sensitive transplant data resides protected in their original sites whilst the comparable small algorithms are exchanged instead. This opens up for the first time the possibility of using health data even across national borders as a basis for developing high-quality clinical prediction models in a protected environment. Health data are of great importance for AI research because they are quality-assured data, which are routinely collected over a long period of time. They are a prerequisite for building reliable prognostic models, but until now they have hardly been accessible for AI research.

In August 2022, representatives of the German partners of the NephroCAGE consortium visited the Canadian cooperation partners. Due to the spread of the COVID-19 pandemic

Members of the German-Canadian NephroCAGE consortium in front of CHUM in Montréal, QC (from left to right: Andreas Schimanski, Klemens Budde, Héloïse Cardinal, Marcel Naik, Verena Graf, Konstantin Pandl, Matthias Niemann, Ruth Sapir-Pichhadze, and Matthieu-P. Shapranow). Photo: M. Schapranow/ NephroCAGE

Members of the NephroCAGE team visiting the HLA laboratory at the McGill University Health Centre – Research Center.
Photo: M. Schapranow / NephroCAGE

Prof. Dr. Klemens Budde giving a lecture to medical trainees at the MUHC-RC.
Photo: M. Schapranow / NephroCAGE

NephroCAGE partners had to collaborate in virtual setup with online tools from the beginning of the project. However, after months of restrictions due to the ongoing COVID-19 pandemic, this was the first opportunity for most project partners to meet in person. Below is a brief summary of the visit and selected agenda items.

The trip begins on Canada's east coast with a visit to Montréal in the Canadian province of Quebec.

Monday, Aug. 15, 2022

NephroCAGE in Montréal, Québec, CA:

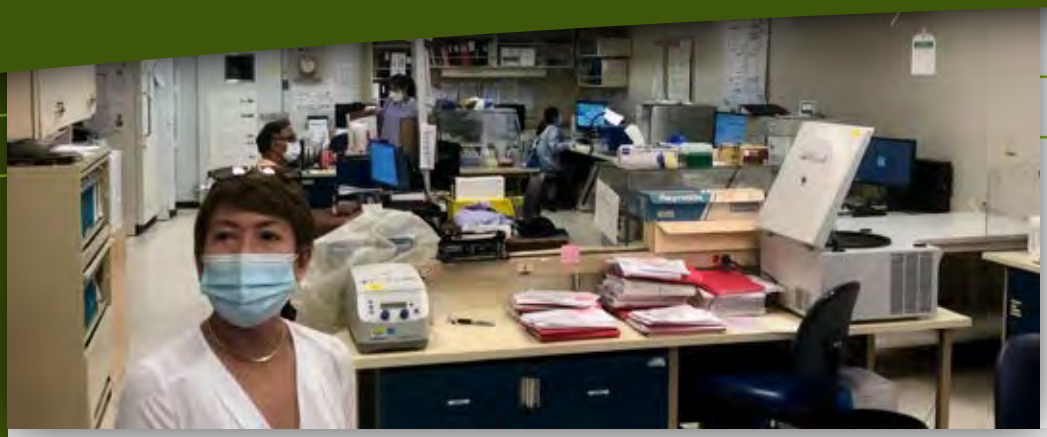
Members of all collaborating partners met for their annual consortium meeting. This internal project meeting served to exchange information on the progress of the individual work packages, upcoming project milestones and overarching project decisions. For this purpose, the NephroCAGE partner centre hospitalier de l'Université de Montréal (CHUM) was kind enough to host us.

Tuesday, Aug. 16, 2022

2nd International NephroCAGE Symposium:

The NephroCAGE consortium presents at international events to report on the latest project results and findings on the application of AI to support nephrologists in kidney transplantation. The International NephroCAGE Symposium was created specifically to facilitate a unique, multidisciplinary exchange among experts from different disciplines, e.g. nephrologists, data scientists, computer scientists and statisticians. The NephroCAGE Symposium started in 2021 - the 50th year of the German-Canadian research partnership - when it gained international attention as a selected event. The 2nd International NephroCAGE Symposium in 2022 was designed in a hybrid format, i.e. participants in Montréal could attend in person while all other participants from all over the world were able to participate via video conference. The program started with greetings from Dr. Matthieu-P. Schapranow, head of the NephroCAGE consortium, and Mr. Kai Sandmann from the German Federal Ministry of Economic Affairs and Climate Action. Two tracks followed: a) Clinical implications from the application of AI on kidney transplantations and b) Impact of digital transformation and application of AI on kidney transplantations. Both tracks contained by short pitch presentations from German and Canadian partners and guests along with subsequent discussion.

We are grateful that many people share the vision of the NephroCAGE team. The number of registered participants increased by more than one third compared to last year and we received overwhelmingly positive feedback from all over the world. In the afternoon, experts from the NephroCAGE teams and participating institutes were invited to attend special workshops covering technical, economic, and regulatory topics. Also on the agenda was a visit to the research center at McGill University Health Centre with a laboratory tour to better understand the existing processes and current data collection in kidney transplantation.



Visit of the HLA laboratory at Vancouver General Hospital.
Photo: M. Schapranow / NephroCAGE

Wednesday, Aug. 17, 2022

Research Deep Dives:

Prof. Dr. Klemens Budde (Charité) shared insights into the latest clinical results from Europe's largest university hospital Charité with medical staff at McGill University Health Centre (MUHC). In addition, the Canadian hosts presented current research projects at CHUM and MUHC to the German visitors.

After visiting the Canadian east coast, the NephroCAGE team continued their journey and traveled to the Canadian west coast to Vancouver in the Canadian province of British Columbia.

Thursday, Aug. 18, 2022

NephroCAGE workshops at the University of British Columbia (UBC):

Together with Canadian partners two days filled with internal workshops followed. Selected topics included harmonization, integration, and analysis of transplant data as well as adaptation of clinical prognostic models to local requirements. In addition, the implementation of the NephroCAGE FLI was an important aspect, because it is designed to ensure training of AI models while maintaining privacy of incorporated data. In addition, participants were allowed to take part in a tour of the UBC campus and its centerpiece – the HLA lab at Vancouver General Hospital. Despite the busy program there was also the opportunity for personal interaction with UBC colleagues.

Contact:

Dr.-Ing. Matthieu-P. Schapranow
Head of the research group "In-Memory Computing for Digital Health" and Scientific Manager Digital Health Innovations,
Hasso Plattner Institute for Digital Engineering gGmbH,
University of Potsdam
schapranow@hpi.de

https://hpi.de/digital-health-center/members/working-group-in-memory-computing-for-digital-health/dr-ing-matthieu-p-schapranow.html

Photo: © private/Schapranow

Members of the NephroCAGE team meet in front of UBC in Vancouver, BC (left to right: Aadil Rasheed, Matthieu-P. Schapranow, Matthias Niemann, Marcel Naik, Andreas Schimanski, Verena Graf, Konstantin Pandl, Sabina Dobrer, Klemens Budde, Karen Sherwood, Paul Keown, Franz Fenninger).
Photo: M. Schapranow / NephroCAGE

events

SBHD 2022

The Music City Spirit Uplifts Scientific Discovery

by Jessica Kimber and Eann Malabanan

Researchers and trainees from around the world brought their expertise and their dancing shoes for the fourteenth annual SYSTEMS BIOLOGY OF HUMAN DISEASE, held entirely in person for the first time in Music City, USA. Vanderbilt University, located minutes from the rich musical history of downtown Nashville, TN, hosted the event in mid-June 2022. After a phenomenal hybrid meeting in Germany in 2021, attendees were excited to collaborate and network in person.

Participants met at the Vanderbilt Commodore Bar and Grille for an opening reception and networking event with live music and local eats. The multifaceted talents of our 90 attendees revealed themselves when one performed an original song during the pre-conference festivities. The vibrant atmosphere of Nashville energized the international community that gathered at the Vanderbilt Student Life Center the following day. While most were able to be together in person, COVID-19 restrictions still hindered many from travelling.

The planning crew of the fourteenth annual SBHD meeting, led by the Vanderbilt University Quantitative System Biology Center (QSBC), the Vanderbilt Diabetes Research and Training Center (DRTC), and the Berlin Institute of Health at Charité (BIH), designed each symposium to be grouped by today's hot topics in human disease research such as Diabetes, Cancer, Aging-related Diseases, and Neurodegenerative Diseases. These topics were chosen with the intention to emphasize their relationships with their COVID-19 research connections. Many distinguished speakers diversified and enriched each of the five symposia, among those were Ioannis Zervantonakis (University of Pittsburgh), Marco Binder (German Cancer Research Center), Alan Attie (University of Wisconsin-Madison), Steven Altschuler (University of California San Francisco), and Yuval Dor (Hebrew University of Jerusalem). Although a few were not able to come to Nashville due to continuing COVID travel

The 2022 Awardee of Anne-Heidenthal Prize for Fluorescence Research, given by Chroma Technology Corp., was Artem Sokolov, PhD from Harvard University for his presentation "Getting the Most out of Highly-multiplexed Images of Tissues and Tumors". His presented work highlights computational tools that are designed for doing end-to-end analysis of highly-multiplexed images. From left to right: Georg Draude (Chroma), Artem Sokolov (Havard), Newell Lessell (CEO Chroma), and Roland Eils (Chair SBHD).
(Copyright: Jessica Kimber)

"The conference was such a fun and edifying experience. It was my first in-person conference, and [it] couldn't have been better."

restrictions, they were still able to engage with about 90 in-person attendees and share their science through a virtual format. In keeping tradition with supporting young researchers, each symposium was designed to highlight both senior and junior investigators in a series of short talks.

Researchers of all levels participated in two poster sessions held throughout the 3-day event to discuss the current trends in systems biology research. Presenters shared hypotheses on cancer therapies and immunology, insulin resistance and obesity, genetic disease risk, and neurodegenerative disease. Selected abstracts were chosen by the planning committee to also give a short, ten-minute talk to summarize their research, along with the opportunity to engage with attendees via Q&A.

The poster sessions provided another opportunity for networking and exchanging ideas among both senior and junior researchers. One attendee said the conference was "such a fun and edifying experience. It was my first in-person conference, and [it] couldn't have been better." Conference attendees further engaged with junior presenters by voting for best poster presentations and short talks.

Two poster sessions offered conference participants the opportunity to present and discuss their work.
(Copyright: Jessica Kimber)

Additional awards were given to four outstanding trainees:

Best Short Talk – $500
Daria Doncevic, Health Data Science Unit, Medical Faculty and BioQuant, Heidelberg
Title: From Black Box to Grey Box - How to improve the biological interpretability of VAE models

Best Short Talk (runner-up) – $250
Rishabh Kapoor, Harvard University
Title: Modeling a tradeoff in IFN-mediated viral defense reveals optimal dynamics and treatment strategies

Best Poster – $300
Julius Upmeier zu Belzen, Berlin Institute of Health, Charité – Universitätsmedizin Berlin
Title: Domain knowledge driven machine learning of genetic disease risk

Best Poster (runner-up) – $150
Samantha Beik, Vanderbilt University
Title: Exploring hypotheses of small cell lung cancer growth mechanisms using Bayesian multimodel inference

Vanderbilt University looks forward to welcoming attendees with the same Music City Spirit and scientific engagement in the Summer of 2024.

Contact:
Jessica Kimber
Program Manager
Vanderbilt Diabetes Center
Nashville
jessica.kimber@vumc.org

events

Hey dude, I see cancer!

EXPLORE Precision Medicine e:Med Event at Futurium Berlin

by Lioba Courth and Silke Argo

Monday morning in Berlin. Classes and MINT courses (MINT: Mathematik, Informatik, Naturwissenschaft und Technik) of four hundred students and teachers stream toward the Futurium. A futuristic building between the main train station and the Spree River, ideal for topics of the future. "EXPLORE Precision Medicine" is the topic of the future today. The special feature: no static exhibition, but the scientists themselves are here and explain their research. A great all-round experience awaits the students: from the offer to try one's hand as a doctor to simulation of the brain and cancer cells under the microscope to VR applications, one can try it all. Interactive exhibition, workshops, lectures, everything at first hand. In addition, there is a booklet with easy-to-understand texts for further study at home and at school. The event is organized by the BMBF-funded research network e:Med Systems Medicine.

"Precision medicine is suitable for dealing with the complex individuality of human beings," emphasizes Prof. Dr. Julie George from the University of Cologne. "With today's spectrum of methods and the expertise that directly benefits the patient in the molecular therapy board, this is the medicine and prevention of the future," explains Prof. Dr. Philip Rosenstiel, UKSH Kiel. The two scientists are spokespersons of the e:Med steering committee.

"We have booked "Learning AI through play"", reports a group of pupils. With simple means, the students themselves become data points and experience an aha effect. The workshop is organized by the young professors Dr. Silke Szymczak, University of Lübeck, and Dr. Helena Zacharias, MH Hannover. "Only now do I realize why our data are so valuable," sums up one student after the GHGA (The German Human Genome-Phenome Archive) workshop on genetic data. Dr. Ulrike Träger and Dr. Julia Philipp from DKFZ Heidelberg conduct it.

"Tumorboard on stage" with Prof. Dr. Angelika Eggert and team, Charité Berlin
(Photo: e:Med, Florian Gaertner, photothek)

Floating through a virtual reality with artificial and artistic organoids
(Photo: e:Med, Florian Gaertner, photothek)

A special highlight is the "Tumorboard on stage" with Prof. Dr. Angelika Eggert and team, Charité Berlin. The pediatric oncologists demonstrate in concrete terms how experts in the clinic decide on an individual therapy. The two hundred students in the hall are fascinated. "What is precision medicine for clinics and research?" give Prof. Dr. Philip Rosenstiel (UKSH Kiel) and Prof. Dr. Joachim Schultze (University of Bonn) insight. "We use AI to determine the right therapy for psychiatric diseases," explains PD Dr. Roberto Goya-Maldonado, University Hospital Göttingen. "Systems medicine helps us track down metabolic diseases in children in particular," says Dr. Annette Bley from UKE Hamburg, explaining the path to diagnosis.

Tour of the exhibition

With the Instagram filter "e:Med Cell-fie," the students become cells themselves. Just behind them, they try out lab coats and pipettes. "Unbelievable, I'm standing in the middle of a ball made of hundreds of molecules!" enthuses one student, immersed in other worlds with VR goggles. Prof. Jörg Menche (University of Vienna) demonstrates this completely new technology for analyzing complex biomedical data. On a giant touchscreen, point clouds light up on and in humans, a project of the Kiel SFB 1182. Undaunted, students explore the microbiome. As DNA detectives, they are on the trail of childhood dementia. Lecturers from the MDC's "Gläsernes Labor" demonstrate an interactive family tree analysis. "I'm totally fascinated by the brain!" can be heard from the next touchscreen. Scientists of the Brain Simulation Section led by Prof. Dr. Petra Ritter, Berlin Institute of Health at Charité (BIH), explain their research on the simulated brain to the students.

Continuing as physician

Students are using transcranial magnetic stimulation to try to non-invasively target brain regions to treat depression, on dummies of course. The project is shown by Dr. Roberto Goya-Maldonado and scientists from AntNeuro. Right next to it, the artistically designed party phone booths of the foundation "Welt der Versuchungen" will beckon. Students get on and enjoy their own music – what do you need to feel good? Here, students are in conversation with addiction researchers around Prof. Rainer Spanagel from the ZI Mannheim. At the Hasso Plattner Institute booth, students dive into the world of IT, technology and software, get information about studying computer science. The online exit game "Escape the Unknown" offers puzzles and knowledge about digital health, personalized medicine and real-world data. Young scientists

Interactive exhibition
(Photo: e:Med, Florian Gaertner, photothek)

Using "Foldscopes", students take pictures of tissue sections with their own cell phones
(Photo: e:Med, Florian Gaertner, photothek)

from Hamburg show examples of how drug repurposing can be used to find customized therapies. In the "Game Zombie Cure Lab," students experience drug repurposing and science computer games with scientists from Erlangen to identify new therapies against melanoma. Behind them are microscopes: "Hey dude I see cancer!" shouts one student. Physician and scientist Dr. Stefan Florian, Charité, shows how to detect tumors and how pathology works today. Scientists from the e:Med project MelBrainSys demonstrate how microscopy techniques help to study skin cancer. Students take great pictures of tissue sections with their cell phones and "Foldscopes." The "Supernoid and Organoid" booth of the Precision Medicine Cluster from Kiel is artistic. Organoids are artificial organ systems for the lab, and students learn how to do research with them. What happens to my data in clinical studies? That's what participants discover at the booth of the "Statusplus" trial portal. "Sequencing on a single cell is ingenious," says one student enthusiastically about single cell sequencing, whose important role in precision medicine is being explained by scientists from Kiel.

Appropriate after the exhibition: young scientists and science slammers explain their complex research in an understandable and entertaining way. Florian Borchert from the Hasso Plattner Institute discusses "Dr. ChatGPT?". Nico Reusch brings along a TÜV for the immune system and thus explains his doctorate at the University of Bonn. Doctoral student Lea Waller, Charité Berlin, explains how genetics helps us to understand the brain.

Exploring something new with what you've just learned

In the career talks, young scientists speak about their path to research. Biologist Dr. Matthia Karreman from Heidelberg University and DKFZ wanted to work in a café when she was a child, but now she is in science. "What made you decide?" she is asked. She smiles "Curiosity!" Now as a computer science professor, I'm at the University of Cologne, Dr. Katarzyna Bozek says, but before that I had jobs across the globe. "What drives you?" is the question. "Exploring something new with what I've just learned," she reveals. Florian Tran, MD, a physician from Kiel, has chosen the path of clinician scientist, which is how he combines clinical practice and research.

"A really great day, but now my head is full," is how one student sums up his condition. By now it's afternoon, groups of students trot out of the Futurium chattering. So many students and teachers have shown their enthusiasm right away. It was definitely worth it to bring precision medicine closer to the students, that is already clear.

e:Med SYSTEMS MEDICINE

Contact:
Dr. Silke Argo
e:Med Management Office
c/o Deutsches Krebsforschungszentrum (DKFZ)
V025 Im Neuenheimer Feld 581
D-69120 Heidelberg
s.argo@dkfz.de

www.sys-med.de
www.sys-med.de/de/explore-precision-medicine

news

EU-STANDS4PM
standards for in silico models for personalised medicine

Normative standards for predictive computer models in personalized medicine

by Marc Kirschner, Project Management Jülich

The EU-Commission funded Horizon2020 project EU-STANDS4PM developed standards for predictive computer-based modeling approaches in personalized medicine. These standards represent an important basis for the harmonization of data integration strategies in medical research and practice, as well as for the further development of normative standardization guidelines for collaborative research projects. This work is thus an important step towards further establishing predictive computer models in personalized medicine and clinical practice in the future.

Two key factors are central to the model building process, for which standardization is important:

- the application of harmonized strategies and methods for the integration of data and
- the validation of models and simulations against the underlying clinical question

Until now there have been very little standards that describe harmonized procedures for data integration and model validation for clinical use. To make a significant contribution in this field, EU-STANDS4PM established a liaison with the *ISO Technical Committee 276 Biotechnology and developed a Technical Specification*[1] as a normative document on an international level. This standard specifies applications of data integration and model validation in personalized medicine. It is encouraging to note that the document has been published on the ISO homepage in June 2023 and is available for purchase. In parallel, a complementary review article *"Computational Models for Clinical Applications in Personalized Medicine – Guidelines and Recommendations for Data Integration and Model Validation"* was published in the Journal of Personalized Medicine[2] that specifically addresses collaborative research projects and presents best practice guidelines for the application of computational models in clinical care.

The future implementation of EU-STANDS4PM's work with ISO will thus play an important role in the development of standardized modeling procedures to improve the quality, reproducibility, and sustainability of research results, as well as to enable easier translation of computational models into clinical practice.

Contact:

Dr. Marc Kirschner
Forschungszentrum Jülich GmbH
Project Management Jülich
Hochschulen, Innovationsstrukturen, Gesundheit (HIG)

Further information at:
www.eu-stands4pm.eu

[1] ISO/TS 9491-1 Biotechnology — Predictive computational models in personalized medicine research — Part 1: Constructing, verifying and validating models

[2] Collin *et al.*, 2022 (PMID: 35207655)

imprint

gesundhyte.de

gesundhyte.de

The magazine for Digital Health in Germany – International Edition Issue 15, December 2023

gesundhyte.de is an annually published magazine with information from German research in the fields of Digital Health and Systems Medicine.

ISSN 2748-3525

This issue is an adapted version of the already published German magazine.

Publisher:

gesundhyte.de is published by the Berlin Institute of Health at Charité (BIH).

Editors:

Editor-in-Chief: Prof. Dr. Roland Eils
Berlin Institute of Health at Charité (BIH)
Editorial Coordination: Franziska Müller
Berlin Institute of Health at Charité (BIH)
Editorial Team:
Dr. Silke Argo (e:Med), Melanie Bergs (PtJ), Prof. Dr. Lars Kaderali (University of Greifswald), Katharina Kalhoff (BIH at Charité), Dr. Marco Leuer (DLR-PT), Dr. Matthieu-P. Schapranow (Hasso Plattner Institute), Dr. Thomas Schmidt (Plattform Lernende Systeme), Dr. Britta Sommersberg (DLR-PT) and Dr. Gesa Terstiege (PtJ).

Address:
Editorial office gesundhyte.de
c/o Berlin Institute of Health at Charité (BIH)
Center of Digital Health
Kapelle-Ufer-2, 10117 Berlin, Germany

Design and layout:
Kai Ludwig, LANGEundPFLANZ digital GmbH, Speyer (www.LPsp.de)

Printed by:
Ottweiler Druckerei und Verlag GmbH, Ottweiler; Germany

Subscriptions:
The magazine is published as part of the public relations work of the publishers. It is supplied free of charge and must not be sold.

For subscription please contact:
Editorial office gesundhyte.de
c/o Berlin Institute of Health at Charité (BIH)
Center of Digital Health
Kapelle-Ufer-2; D-10117 Berlin
franziska.mueller@bih-charite.de

apops - Fotolia

Please visit our homepage if you would like to find out more about the topics of our magazine.

What you can expect:

- Exciting stories from everyday research – find out more about ongoing projects
- Interviews and profiles – get to know the faces behind the research
- Extensive overview of events – never miss another important date
- Information about current funding – stay up to date
- Get involved – suggest your own topic

We look forward to your visit on our homepage!

about us

Presenting the gesundhyte.de editorial team

gesundhyte.de aims to communicate the successes of the German research to a broad audience in a descriptive way. The magazine is created once a year in German and English by a multidisciplinary editorial team from various German research institutions: Berlin Institute of Health at Charité (BIH), Hasso Plattner Institute Potsdam, University of Greifswald, Project Management Jülich, DLR Project Management and representatives of the initiatives: Lernende Systeme – the Platform for Artificial Intelligence in Germany and e:Med Systems Medicine. The magazine is financed by the Berlin Institute of Health at Charité (BIH).

Front row, left to right: Prof. Dr. Roland Eils (BIH at Charité), Franziska Müller (BIH at Charité), Dr. Silke Argo (e:Med), Melanie Bergs (PtJ), Prof. Dr. Lars Kaderali (University of Greifswald),
second row, left to right: Katharina Kalhoff (BIH at Charité), Dr. Marco Leuer (DLR-PT), Dr. Matthieu-P. Schapranow (Hasso Plattner Institute), Dr. Thomas Schmidt (Plattform Lernende Systeme),
third row, left to right: Dr. Britta Sommersberg (DLR-PT), Dr. Gesa Terstiege (PtJ), and Kai Ludwig (LANGEundPFLANZ digital GmbH)

Photos: shutterstock_© RAY-BON / Eils, © Sebastian Runge / Müller, private / Argo, private / Bergs, private / Kaderali, private / Kalhoff, © Thomas Rafalzyk / Leuer, private / Schapranow, private / Schmidt, © Thilo Schoch / Sommersberg, private / Terstiege, private / Ludwig, private

contact

Berlin Institute of Health at Charité (BIH)
Editor in Chief: Prof. Dr. Roland Eils
Editorial coordination and subscription services: Franziska Müller
Editorial assistance: Tatjana Schütz und Vivien Schubert
c/o Berlin Institute of Health at Charité (BIH)
Center of Digital Health
Kapelle-Ufer-2; D-10117 Berlin, Germany
E-Mail: franziska.mueller@bih-charite.de
www.hidih.org

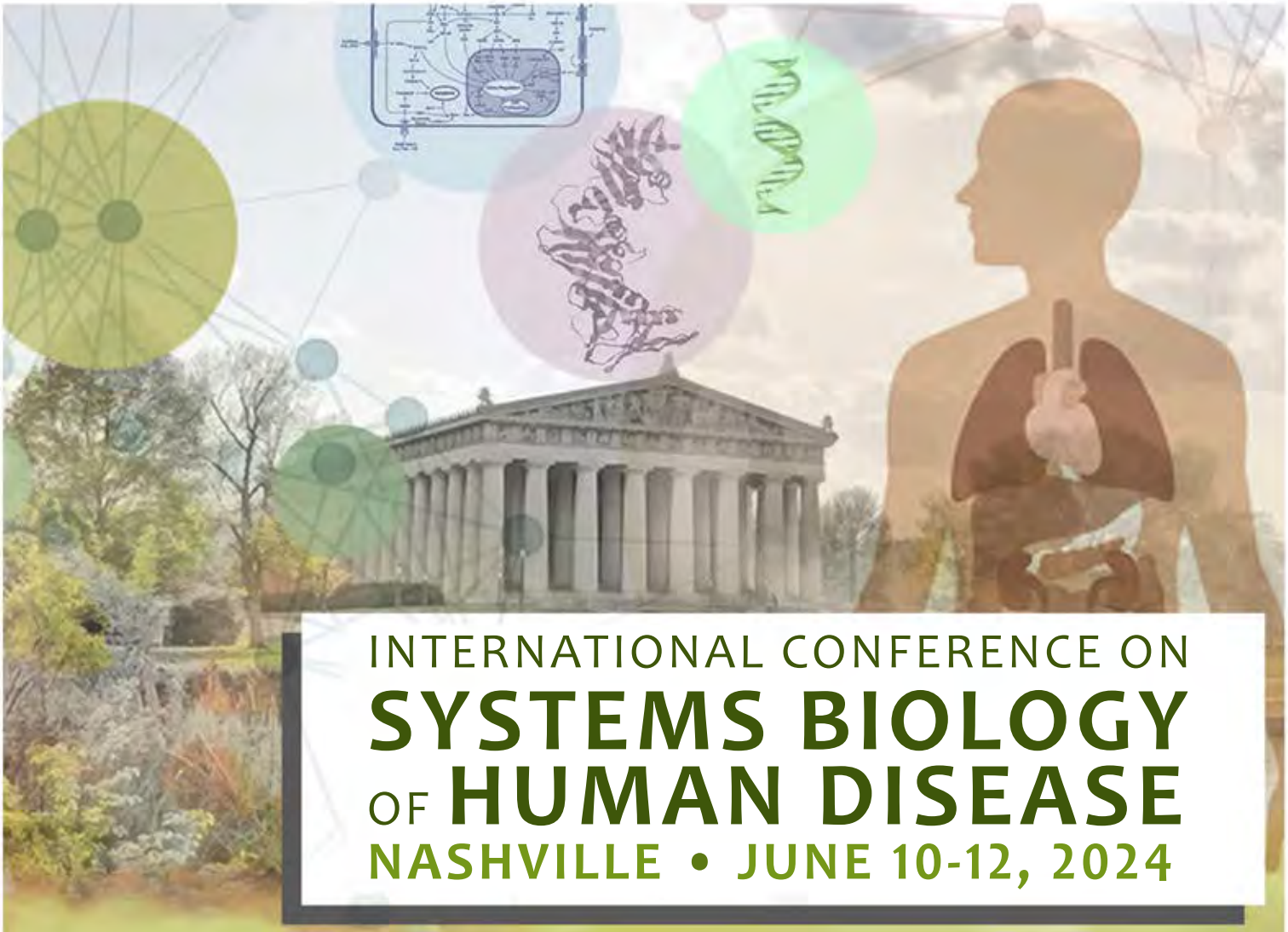
Communication and Marketing, Press Office
Contact person:
Katharina Kalhoff
Anna-Louisa-Karsch-Str. 2; D-10178 Berlin, Germany
E-Mail: katharina.kalhoff@bih-charite.de
www.bihealth.org

Lernende Systeme – Germany's Platform for Artificial Intelligence
Contact person:
Dr. Thomas Schmidt
Managing Office: c/o acatech
Pariser Platz 4a; D-10117 Berlin, Germany
E-Mail: t.schmidt@acatech.de
www.plattform-lernende-systeme.de

Project Management Jülich
Forschungszentrum Jülich GmbH
Universities, Innovation Structures and Health (HIG)
Contact persons:
Dr. Gesa Terstiege and Melanie Bergs
D-52425 Jülich, Germany
E-Mail: g.terstiege@fz-juelich.de, m.bergs@fz-juelich.de
www.ptj.de/en

German Aerospace Center (DLR)
Project Management Agency
Department Health (OE20)
Contact persons:
Dr. Marco Leuer, Dr. Bettina Koblenz and Dr. Britta Sommersberg
Heinrich-Konen-Str. 1; D-53227 Bonn, Germany
E-Mail: marco.leuer@dlr.de, bettina.koblenz@dlr.de and
britta.sommersberg@dlr.de
www.dlr-pt.de

e:Med Management Office
Managing Director:
Dr. Silke Argo
c/o Deutsches Krebsforschungszentrum (DKFZ) – V025
Im Neuenheimer Feld 581; D-69120 Heidelberg, Germany
E-Mail: s.argo@dkfz.de
www.sys-med.de



INTERNATIONAL CONFERENCE ON SYSTEMS BIOLOGY OF HUMAN DISEASE

NASHVILLE • JUNE 10-12, 2024

www.sbhdNashville.org

SAVE THE DATE

SBHD USA at Nashville, TN

ORGANIZED BY:

ROLAND EILS
BERLIN INSTITUTE OF HEALTH AT CHARITÉ – UNIVERSITÄTSMEDIZIN BERLIN, GERMANY

IRINA LEHMANN
BERLIN INSTITUTE OF HEALTH AT CHARITÉ – UNIVERSITÄTSMEDIZIN BERLIN, GERMANY

AL POWERS
VANDERBILT UNIVERSITY SCHOOL OF MEDICINE, NASHVILLE, USA

VITO QUARANTA
VANDERBILT UNIVERSITY, NASHVILLE, USA

SUPPORTED BY:

BIH Berlin Institute of Health @Charité

VANDERBILT UNIVERSITY

/vu/qsbc quantitative systems biology center

VANDERBILT UNIVERSITY MEDICAL CENTER

CHROMA®