

---

# **Abstracts**

## **Matchmaking WASP - DDLs**

Joint call 2021

---

*UPDATED: August 24, at 15:00*

[Information about the call @ SciLifeLab webpage](#)  
[Information about the call @ WASP webpage](#)

**WASP DDLS joint working group**

Bo Bernhardsson (LU)- WASP  
Ingrid Hotz (LiU) - WASP  
Joakim Jaldén (KTH)-WASP  
Rebecka Jörnsten (CTH/GU)- WASP  
Matts Karlsson (LiU)- DDLS  
Danica Kragic (KTH)- WASP  
Erik Kristiansson (CTH)- DDLS  
Erik Lindahl (SU)-DDLS  
Amy Loutfi (ÖrU)- WASP  
Martin Rosvall (UmU)- WASP  
Carolina Wählby (UU)- DDLS  
Anders Ynnerman (LiU)- WASP

**Support and questions**

ddls-calls@scilifelab.se

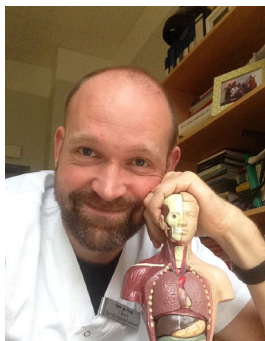
# Abstracts

## Matchmaking WASP - DDLS

Find your collaborator for the joint call 2021

*Abstracts listed in incoming order*

- |                         |                              |
|-------------------------|------------------------------|
| 1. Cornelis Jan Pronk   | 19. Tuuli Lappalainen        |
| 2. Torbjörn Lundh       | 20. Maria Lerm               |
| 3. Ahmed Ali-Eldin (1)  | 21. Tanja Slotte             |
| 4. Ahmed Ali-Eldin (2)  | 22. Simon Elsässer           |
| 5. Amir Aminfar         | 23. Kevin Smith              |
| 6. Oleg Sysoev          | 24. Björn Wallner            |
| 7. Oscar Martinez Mozos | 25. Andreas Kerren           |
| 8. Mathias Uhlen        | 26. Maria Sandsten           |
| 9. Saeed Shoaie         | 27. Stefan Bauer             |
| 10. Cheng Zhang (1)     | 28. Linn Fagerberg           |
| 11. Cheng Zhang (2)     | 29. Pontus Nordenfelt        |
| 12. Raazesh Sainudiin   | 30. Erik Svensson            |
| 13. Fredrik Edfors      | 31. Maria Fällman            |
| 14. Sonja Aits          | 32. Kristoffer Sahlin        |
| 15. Patrick Lambrix     | 33. Anne-Marie Fors Connolly |
| 16. Henrik Green        | 34. Paul Hudson              |
| 17. Ola Spjuth          | 35. Mohammad Ghorbani        |
| 18. Removed             |                              |



## Cornelis Jan Pronk

Community: Life Science  
Senior consultant pediatric oncology,  
Associate Professor  
kees-jan.pronk@med.lu.se



**Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it**

**Research areas: precision medicine and diagnostics, epidemiology and infection biology**

**Key words: hemorrhagic cystitis, BK-virus, genetic polymorphisms, biomarkers, hematopoietic stem cell transplantation**

Models to predict clinical hemorrhagic cystitis after hematopoietic cell transplantation by integrating clinical phenotypes, genetic polymorphisms, urinary biomarkers and BK-virus titres

BK virus-associated hemorrhagic cystitis (HC) is a debilitating complication following allogeneic hematopoietic cell transplantation (allo-HCT). Conditioning, irradiation, HLA-disparity and T-cell depletion have been identified as risk factors of HC. However, knowledge of biomarkers predictive of HC are scarce. Also, the role of BK virus (BKV) in HC remains unclear, as illustrated by our previously study in which only 50% of patients with urinary BKV developed HC in a cohort of pediatric allo-HCT. This latter observation may perhaps be explained by genetic factors/polymorphisms that increase susceptibility to develop clinical HC in these patients. This suggestion has support in literature as genetic factors have been implicated in the development of urinary tract infections in other contexts.

The aim to increase our knowledge on BKV-associated HC, with the goal to decrease the risk for this complication and reduce or ameliorate the suffering it causes. If a role of BKV levels predicting severity of the HC can be established, BKV-specific T-cell treatment could become an integral part of clinical practice. Further, if genetic polymorphisms or biomarkers could predict development prior to clinical presentation, actions to prevent or ameliorate the symptoms could be made in the clinic.

In a multicenter setting, involving all pediatric HCT units in Sweden (4 total), and the adult HCT in Lund, we aim to prospectively study the incidence of BKV and HC in a large cohort of pediatric and adult patients. During a 3-year period, approximately 150 children and 150 adults are expected to participate. For all patients, diverse clinical data will be collected, including age, sex, underlying disease, transplantation and conditioning modality, immunological recovery as well as development of HC. In addition, urinary samples on all of these patients will be screened for BK-virus levels, as well as levels of approximately 12 urinary markers including inflammatory cytokines and kidney

injury markers. Lastly, for most adults and a part of children, whole genome sequencing (WGS) is planned to allow us to perform wide GWAS studies.

In my own research group, we aim to perform "single level" descriptive analysis/descriptions of the finding in this study for all of the data, excluding the WGS data. We would however be interested to collaborate with a group that has the expertise to do integral analysis of all of these data, and to investigate if findings in the the different "datasets" (i.e. 1) phenotypic/clinical data, 2) BK-levels, 3) urinary biomarker levels, 4) WGS) interact or associate. The ultimate goal would be if we could built a model that predicts development of HC prior to clinical presentation. I think previous experience with analysis of host-genome data (incl GWAS) is important. As the project has not started yet, there still is the possibility to discuss with the collaborative partner if the type of data and/or data-collection itself could be optimised for analyses purposes. Patient-inclusion is planned to start late 2021-early 2022. If the project would be granted, part of it would go to generation of urinary marker and WGS data for some of the patients. I am open discuss how collaboration, and the type of application, would be most applicable.

### **Short Bio**

I moved from Netherlands to Sweden 2002 as an M.D.. Since then I have worked clinically 50% and research 50%. My PhD-studies were preclinical models to study hematopoietic stem cell biology. Clinically, I am pediatrician and trained pediatric oncologist with a focus on leukemia and hematopoietic cell transplantation (HCT). Currently, I have preclinical and clinical research projects ongoing, all with a focus on leukemia and HCT. I am a clinical fellow in Hematology within the Wallenberg Centre Molecular Medicine in Lund.



## Torbjörn Lundh

Community: WASP

Professor in Biomathematics, Chalmers

torbjorn.lundh@chalmers.se



### **Looking for open problems/questions to create new collaborations**

### **Research areas: evolution and biodiversity, AI/Math**

**Key words: problem based, needs finding, mathematical models, agent based, differential equations, parameter fitting, data augmentation**

## Mathematical-Meta-Matchmaking

Models are central in all science today, and mathematics seems, according to Wigner, to be an "[unreasonable effective](#)" way to treat scientific models. See the illustration of the various model classes in use today [1].

In Life Science, we see that new methods to collect massive amount of information, puts the data in the driving seat. This will lead to a golden opportunity to use various viewpoints, i.e. models, to gain understanding of the many complex phenomena in that field.

I would like to take this match-making opportunity to help Life Science researchers to find new ways to put mathematics in an "unreasonable" use. Not by myself, but with other mathematicians more skilled in the specialized tools that might attack the phenomena at hand. That's why I created the database [mathbio.se](#) in order to facilitate interdisciplinary matching in Gothenburg between Life Science researchers and mathematicians, in an extended meaning.

Let me finish with a more recent example from the AI field to illustrate what I would like to help out with. [Andrew Ng](#) gave a keynote presentation at the med-school in Stanford where he stated that the main future challenges to use AI for clinical problems was lack of data. He pointed out two ways to attack that challenge: data augmentation and more clever architecture. The latter point can be viewed as a need for better models. Regarding the first challenge, I contacted his group with a little idea that resulted in [2]. When they then started to talk about a topological problem, I felt that it would be better to direct them to the mathematical nestor in that field, Gunnar Carlsson, that recently resulted in [3].

Please contact me if you got a problem you would like to unconditionally discuss in order to add other possible viewpoints and new analytic tools.

1. Gerlee, Philip, and Torbjörn Lundh. [Scientific models: red atoms, white lies and black boxes in a yellow book](#). Springer, 2016.
2. Zhou, Sharon, Jiequan Zhang, Hang Jiang, Torbjörn Lundh, and Andrew Y. Ng. "[Data augmentation with Mobius transformations](#)." Machine Learning: Science and Technology 2, no. 2 (2021): 025016.
3. Zhou S, Zelikman E, Lu F, Ng AY, Carlsson G, Ermon S. [Evaluating the Disentanglement of Deep Generative Models through Manifold Topology](#). ICLR 2021.

## Short Bio

Torbjörn Lundh, professor in biomathematics at Chalmers.  
Assistant Dean, Faculty of Science, University of Gothenburg  
Member in WASP's Graduate School Management

MSc, Engineering Physics, Uppsala, 1990

PhD, Mathematics, Uppsala, 1996

Lecturer at KTH, 1996

Three postdocs: Cambridge, UK; Stony Brook, NY, USA and Institute Mittag-Leffler, Djursholm

Sabbatical: Visiting professor at Vascular Surgery, Stanford, 2015-2016

Previously:

Chairman of the [Swedish National Committee for Mathematics](#) under KVA

Vice president of [European Society for Mathematical and Theoretical Biology](#)

Co-founder and co-inventor of four medtech spin-out companies: [y-Graft](#), [GraftCraft](#), [PressCise](#) and [Navari AB](#)

Mail: [torbjorn.lundh@chalmers.se](mailto:torbjorn.lundh@chalmers.se)

Tel: 0709847070

[Home page](#)

## Ahmed Ali-Eldin

Community: WASP  
Asst. Prof. Chalmers  
ahmed.hassan@chalmers.se



**Looking for open problems/questions to create new collaborations**  
**Have an open project or idea that could benefit from a collaboration**

**Key words: Medical devices; Neuroscience; Sensors**

### Edge Clouds for Medical devices

In a recent discussion with colleagues working on neuroscience from the Netherlands, one interesting problem that we discussed was the usage of edge cloud resources for medical implants and devices, including, Neuro-Prosthetics for vision, where the implant needs to be very low energy, with very little computing power, but can make use of edge cloud resources to run DNNs for vision. In general, there are many possibilities for using edge computing for health-care, and rehabilitation assistive device. This direction of work can take multiple different directions based on the joint competence we can have in the project. Some exemplary work:

1. Otseidu, Kofi, et al. "Design and optimization of edge computing distributed neural processor for biomedical rehabilitation with sensor fusion." Proceedings of the International Conference on Computer-Aided Design. 2018.
2. Wang, Xiaying, et al. "An Accurate EEGNet-based Motor-Imagery Brain-Computer Interface for Low-Power Edge Computing." 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA). IEEE, 2020.

Will be happy to discuss these different directions.

### Short Bio

I am an asst. Prof, at Chalmers. Before that I was a postdoc at UMass Amherst for three years. My main competence is in distributed systems, cloud computing, edge cloud computing, and ML systems.



## Ahmed Ali-Eldin

Community: WASP

Asst. Prof. Chalmers

ahmed.hassan@chalmers.se

4

**Looking for open problems/questions to create new collaborations  
Have an open project or idea that could benefit from a collaboration**

**Research areas: AI/MLX, Autonomous systems, Software**

**Key words: ML; AI; Systems; Machine Learning**

### ML Systems for Medicine

Most Machine learning systems have been developed for Internet scale companies (Google, Facebook, Uber), with very little work on systems developed for life-sciences. Many researchers have argued that in their current state, ML systems are sub-optimal [1] and still requires quite a bit of work. In addition, while having general ML systems such as PyTorch and Tensorflow results in ease of adoption, it also leads to reduced performance and efficiency for many applications. Hence, I believe taking a clean-slate design approach to these systems that starts from an application, in this case life-sciences, and re-architecting ML systems to optimize their performance and efficiency will probably have a very high-impact on what can be achieved. I am thus interested in collaborating with researchers in the life-science domains on this journey of truly harnessing the power of ML systems.

[1] Barham, Paul, and Michael Isard. "Machine learning systems are stuck in a rut." Proceedings of the Workshop on Hot Topics in Operating Systems. 2019.

### Short Bio

I am an asst. professor at Chalmers Network and Computer Systems division. My research focus has been on systems at large, and distributed systems including edge and cloud computing. Prior to coming to Chalmers, I spent three yrs at UMass Amherst where I am still affiliated.

## Amir Aminifar

Community: WASP

Associate Senior Lecturer, Lund University

amir.aminifar@eit.lth.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: precision medicine and diagnostics, AI/MLX**

**Key words: Trust, Robustness, Machine Learning, Precision Medicine and  
Diagnostics**

### Trustworthy Machine Learning in Precision Medicine and Diagnostics

Most Machine learning systems have been developed for Internet scale companies (Google, Facebook, Artificial Intelligence (AI) has attracted a lot of attention in the past few decades. Trust in AI, however, represents a major challenge, particularly when it comes to precision medicine and diagnostics. An interesting research direction is to investigate the potential risks associated with the adoption of AI techniques in such domains and provide support for trust in the decisions made by the AI and machine learning techniques.

#### Short Bio

Amir Aminifar is a WASP Assistant Professor (Biträdande Universitetslektor) in the Department of Electrical and Information Technology at Lund University, Sweden. He received his Ph.D. degree from the Swedish National Computer Science Graduate School (CUGS), Linköping University, Sweden. During 2016-2020, he held a Scientist position in the Institute of Electrical Engineering at the Swiss Federal Institute of Technology (EPFL), Switzerland. His research interests include health informatics and machine learning in precision medicine and diagnostics, as well as wearable and m-health.



## Oleg Sysoev

Community: WASP  
Associate Professor  
oleg.sysoev@liu.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration**

**Research areas: cell and molecular biology, precision medicine and diagnostics,  
AI/MLX, Autonomous systems**

**Key words: personalized medicine, subgroup identification, decision trees,  
explainable AI, deep learning**

### Machine Learning for personalized medicine

My primary research interest is development of novel machine learning methods for different applications. Related to the WASP-DDLS area, I am currently working with development of new methods for analysis of single cell data (in particular, novel methods for cell typing), and I have some ideas regarding extending my work towards individualized drug prediction and drug repurposing. I am also interested in development of new methods for data integration and transfer learning for omics data. I have been also working with public health applications and developed a novel decision tree method for personalized medicine based on clinical data.

I am interested to collaborate with a person from the DDLS community having good knowledge of applied aspects and, more importantly, genuine interest to create novel machine learning tools for personalized medicine rather than testing existing machine learning methods for solving applied problems. I am not bound to the methodologies that I am currently using and I am open to explore new methods and ideas (given that they are related to methodological development)

### Short Bio

I am associate professor (docent) in Statistics at Linköping University working with machine learning methods and applications. I am also the head of the master's programme in Statistics and Machine Learning at LiU, where I teach many courses, a course in Machine Learning being one of them.

In addition to personalized medicine related methods, I have been developing methods for large-scale monotonic regression, machine learning for telecommunication (in collaboration with WASP and Ericsson), machine learning for public health (primarily decision trees), statistical methodology (primarily bootstrap methods).

## Oscar Martinez Mozos

Community: WASP  
Universitetslektor (Docent),  
Örebro University  
omozos@gmail.com



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: precision medicine and diagnostics, AI/MLX, Autonomous systems,  
Software**

**Key words: mental well-being, artificial Intelligence, machine learning, wearable  
sensors, physiological signals, daily life, AAL sensors, psychology**

### AI for mental well-being monitoring and assessment during daily life

Most Machine learning systems have been developed for Internet scale companies (Google, Mental well-being is one of the main concerns in the society. Artificial Intelligence can support the monitoring and assessment of mental well-being during daily life activities by analysing different types of data like physiological or activity signals, from different sources like wearables, smartphones, AAL sensors, or short questionnaires. Important challenges include the collection of long-term datasets for training the algorithms, the inclusion of multiple modalities, and the personalisation of the system for each user.

#### Short Bio

I am Universitetslektor (Docent), and WASP-AI faculty member. I have a background on AI, machine learning, robotics and autonomous systems, neuroscience, and psychotherapy.

My homepage: <https://sites.google.com/view/oscarmozos>



## Mathias Uhlen

Community: Life Science  
Professor, KTH/KI  
mathias.uhlen@scilifelab.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it**

**Research areas: cell and molecular biology, precision medicine and diagnostics,  
AI/MLX**

**Key words: Data integration, protein atlas, precision medicine**

### The Human Protein Atlas

The Human Protein Atlas (HPA) program aims to map of all the human proteins in cells, tissues and organs using integration of various omics technologies, including proteomics, transcriptomics, antibody-based methods and AI-based systems biology. The current version ([www.proteinatlas.org](http://www.proteinatlas.org)) consists of separate parts and the database contains in total more than 20 million web pages with information on various aspect of the genome-wide analysis of the human proteins. All the data in the knowledge resource is open access to allow scientists both in academia and industry to freely access the data for exploration of the human proteome. We welcome all efforts within the DDLS program to expand the value of this SciLifeLab-based resource.

Selected and recent publications:

- Uhlen et al (2015) "Tissue-based map of the human proteome" Science 347(6220):1260419.  
Uhlen et al (2017) "A pathology atlas of the human cancer transcriptome" Science 357(6352): eaan2507  
Uhlen et al (2019) "A genome-wide transcriptomic analysis of protein-coding genes in human blood cells" Science366 (6472): eann9198  
Sjöstedt et al (2020) "An atlas of the human, pig and mouse brain" Science 367 eaay5947  
Mahdessian et al (2021) "Spatiotemporal dissection of the cell cycle with single-cell proteogenomics" Nature 590 (7847): 649-654.  
Karlsson et al (2021) "A single cell type transcriptomics map of human tissues", Science Advances, in press.

## Short Bio

Mathias Uhlen research is focused on protein science, precision medicine and data-driven systems biology. His research has resulted in more than 650 publications with an h-index of 123 (Google Scholar). Since 2003, he has led an international effort to systematically map the human proteins to create a Human Protein Atlas using omics technologies. He is a member of the National Academy of Engineering (US), The Royal Swedish Academy of Science (KVA), the Royal Academy of Engineering Sciences (IVA) and European Molecular Biology Organization (EMBO). From 2010-2015, he was the Founding Director of SciLifeLab.

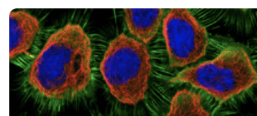
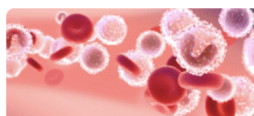
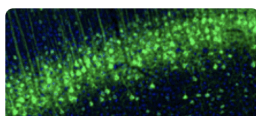
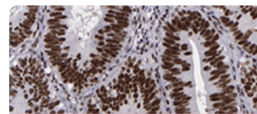
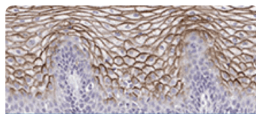
# THE HUMAN PROTEIN ATLAS

≡ MENU    HELP    NEWS

SEARCH<sup>1</sup>

Search [Fields »](#)

e.g. ACE2, GFAP, EGFR





## Saeed Shoaie

Community: Life Science  
Group leader, KTH  
saeed.shoaie@scilifelab.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: precision medicine and diagnostics, AI/MLX**

**Key words: Microbiome, Systems Biology, Microbial Metabolism, Metagenomics,  
Antimicrobial Resistance, Network biology**

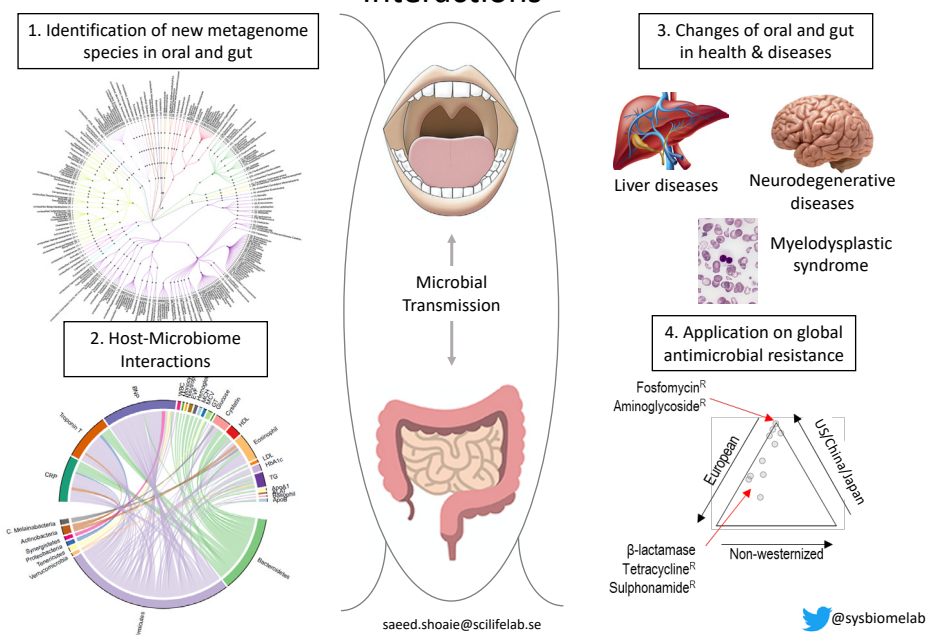
Human Gut Microbiome Atlas; Pan-metagenomics analysis on compositional and functional changes of human gut microbiome in health and disease

The human gut microbiota in humans associated to several diseases and provided the possibilities for analyzing microbial richness in health and disease, probing factors such as geographic and environmental differences. We investigated the gut microbial changes of human populations across 19 countries and 20 different diseases to explore the compositional and functional changes. Our analyses, involving 5,224 samples, included a large number of data from patients with diseases such as diabetes, obesity, fatty liver disease, and cancer, allowing comparison across both geography and disease profiles. Additionally, to study the temporal changes of the human gut microbiome, we followed 86 healthy individuals over one year, with four sampling times and to analyse the stability of the microbiome during one year. The integrative analysis of longitudinal microbiome changes revealed the existence of two types of species with variation in abundance with that one type tends to be enriched in disease, and the other type is enriched in health. We suggest that the decrease type one species and the increase of type two species with time could be unrecognized aspect of the homeostasis maintenance in a healthy microbiome. In addition, we developed a comprehensive computational platform for personalized and community metabolic modelling of microbiome. Based on our modelling approach, we suggested a new the stratification concept, reactobiome, based on the metabolic features of the human gut microbiome. We developed the reactobiome for healthy population across 16 countries and five reactotypes were observed with specific amino acid, carbohydrate and xenobiotic metabolic features.

## Short Bio

We investigate the global and temporal changes of human microbiome in health and diseases. We develop computational and quantitative systems biology tools and models to understand the role of microbiome in host physiology and ultimately propose an improved microbiome-based personalized medicine pathway for future therapy. We work on different microbial gene catalogues together with metagenome species, including bacteria, fungi and phage to achieve better representation of uncultured genomes in the human microbiome. We characterize the metagenome species and gene catalogues with several functional annotations such as antimicrobial resistance, virulence factors, mobile genetic elements, and secondary metabolites. We then apply this information to understand the distinct and shared microbial functions on different body sites and several diseases. In different studies, we analyse multi-omics data on host and microbiome. We integrate these data in a personalized fashion using integrated network system. The outcomes will lead to a personalized approach and better stratification of patients based on multi-dimensional data integration. These findings will also provide valuable information to perform personalized predictive modelling to simulate and increase treatment efficacy based on host-microbiome interactions.

## Translational Systems Biology in Host-Microbiome Interactions







## Cheng Zhang

Community: Life Science  
Researcher, KTH  
cheng.zhang@scilifelab.se



**Have an open project or idea that could benefit from a collaboration,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: cell and molecular biology, precision medicine and diagnostics**

**Key words: Whole body modeling, systems biology, genome-scale metabolic modeling, co-expression network, protein-protein interaction network**

### Development of whole body integrative model to facilitate integration of interdisciplinary data for Biomarker and Drug Development

Systems biology tools, such as genome-scale metabolic models (GEMs), protein-protein interaction networks (PPINs) and co-expression networks (CNs), has been powerful tools for studies in many diseases including, obesity, Non-Alcoholic Fatty Liver Disease (NAFLD) and cancers. In my recent project, the pathology atlas, which has been published on Science in August this year, we demonstrated that the use of systems biology tools, namely cancer specific CNs and patient specific GEMs, can help us to see the big picture of different cancers.

Although the current systems biology platforms cover the link among metabolites, proteins and genes within a specific cell/tissue, the links are often binary which limited their applications. In addition, human bodies are extremely complicated systems, and different organs are synchronized with each other in a highly organized and systematic way. So far, there is no mature whole body scale model that could handle this complexity.

Here, the aim is to setup a novel platform, namely whole body multi-scale GEM (WBGEM), which facilitate for studies of human wellness and diseases by integrating multi-level biological networks into GEMs. The prediction of this platform will be tested in human cell line case and validated with in vitro experiments. In addition, tissue specific multi-scale GEMs will be built using this platform and combined into a WBGEM. This WBGEM will be used to predict the whole body response of gene knock out and validated with mouse experiment. Moreover, the WBGEM for human will be used in a real case to investigate the underlined biological insight of a disease, e.g. the connection between plasma mannose level and T2D as well as cardiovascular disease (CVD), in a whole body scale.

The project is of strategic importance to studies in human diseases and pharmaceutical industry: WBGEM developed in this study is expected to assist on: i) systematic understanding of development and progression of human diseases; ii) identification of novel drug targets and biomarkers, iii) improving insights into the modes of action of existing drugs, iv) enabling unsuccessful drug development to be terminated before it becomes too costly, v) facilitating early diagnosis, to shift from disease treatment to disease prevention vi) enabling individualized prescriptions with personalized input data, to avoid the prescription of drugs to non-responders or to patients likely suffering from severe side effects.

### **Short Bio**

Dr. Cheng Zhang is an experienced systems biologist, and he has conducted the creation of pathology atlas and single cell type atlas in within the Swedish Human Protein Atlas program. His main research interests including developing and expanding genome-scale metabolic models to other omics level, and applying them into clinic such as investigating the driving factors, biomarkers or potential therapeutic strategies for obesity, non-alcoholic fatty liver disease and hepatocellular carcinoma.



## Cheng Zhang

Community: Life Science  
Researcher, KTH  
cheng.zhang@scilifelab.se



**Looking for open problems/questions to create new collaborations,  
Have a dataset and want to see if there is more that can be done with it,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: cell and molecular biology, precision medicine and diagnostics**

**Key words: Single cell sequencing, systems biology, bioinformatics, human protein atlas, transcriptomics, proteomics**

### A single cell type transcriptomics map of human tissues

Advances in molecular profiling have opened up the possibility to map the expression of all genes in all cells, tissues and organs in the human body. Here, we combined single cell transcriptomics analysis with spatial antibody-based protein profiling to create a high-resolution single cell type map of human tissues. An open access atlas has been launched in order to allow researchers to explore the expression of all human protein-coding genes in 192 individual cell type clusters. An expression specificity classification was performed to determine the number of genes elevated in each cell type allowing comparisons with bulk transcriptomics data. The analysis highlights distinct expression clusters corresponding to cell types sharing similar functions, both within the same organs and between organs.

We plan to update the Single Cell Type Atlas in the future as part of the annual update of the Human Protein Atlas taking into account new genome assemblies and addition of new data. Future inclusion of more tissues and datasets with more cells analyzed and sequencing at higher depths will allow this effort to be extended into more specialized tissues and organs helped by international efforts such as the Human Cell Atlas and other efforts to analyze human single cells in tissues. The analysis may also be extended to include single-nuclei data from tissues that are difficult to obtain using the single cell approach. Integration of these datasets both on the transcriptomic and proteomic level gives a unique opportunity to validate the exact tissue localization in situ. Thus, the approach described here to combine single cell data with antibody-based profiling will facilitate efforts to provide a holistic cell-, tissue- and organ-wide map of the proteins across the human body to act as a basis for research in human biology and disease.

## **Short Bio**

Dr. Cheng Zhang is an experienced systems biologist, and he has conducted the creation of human pathology atlas and single cell type atlas within the Swedish Human Protein Atlas program. His main research interests including developing and expanding genome-scale metabolic models to other omics level, and applying them into clinic such as investigating the driving factors, biomarkers or potential therapeutic strategies for obesity, non-alcoholic fatty liver disease and hepatocellular carcinoma. His research resulted in the development of a few novel drugs for fatty liver disease and different cancer types.



## Raazesh Sainudiin

Community: WASP  
Associate Professor of Mathematics  
with Specialisation in Data Science,  
Department of Mathematics,  
Uppsala University  
raazesh.sainudiin@math.uu.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration**

**Research areas: evolution and biodiversity, epidemiology and infection biology, AI/  
MLX, AI/Math, Software**

**Key words: petabyte-scale nonparametric density estimators, tree arithmetic for  
statistical operations, feature engineering for decision-support pipelines**

### Scalable Statistical Regular Pavings

I am mainly looking for concrete problems involving massive amounts of data from the life-sciences that require custom-built mathematical and statistical models with concomitant software engineering to go from data to decisions in appropriate computing infrastructure. Specifically, recent work is underway on scalable nonparametric density estimators and other more generic arithmetic operators that can aid in statistical decision-making, including classical and Bayesian hypothesis testing, estimation for model selection and/or for predictive or control purposes. The density estimators can work or be trained jointly with specific deep neural architectures, depending on the problem domain, and may be widely seen as a formal way to introduce classical scientific decision problems into over-parametrized and appropriately regularized AI models. Moreover, the underlying data can be time-series of real-valued vectors, matrices, tensors, images, 3D reconstructions of shapes, among other measurable events. The tools are open by design.

Let's have a chat if you are interested. See the following: <https://doi.org/10.1007/s42081-019-00054-y> and <https://arxiv.org/abs/2012.14847>. For track-record of successful collaborations with life scientists in the past see: <https://doi.org/10.1016/j.tpb.2018.07.002>, <http://dx.doi.org/10.1007/s00285-015-0886-z> or <http://www.genetics.org/cgi/content/abstract/168/1/383> and <http://dx.doi.org/10.1016/j.jtbi.2016.07.038>.

### Short Bio

I, Raazesh Sainudiin, work at the interface of computing, mathematics and statistics to solve real-world problems through custom-built mathematical and statistical models. I completed my PhD in Mathematical Statistics at Cornell University, a postdoctoral fellowship at Department of Statistics, Oxford University, Senior Lectureship in Statistics at University of Canterbury, New Zealand. I currently have a joint academic-industrial appointment at Department of Mathematics, Uppsala University and Combient Mix AB, Stockholm.



## Fredrik Edfors

Community: Life Science  
Researcher, SciLifeLab, KTH  
fredrik.edfors@scilifelab.se



**Have a research tool or approach and want to find a collaborator with an idea or data to apply it to**

**Research areas: precision medicine and diagnostics**

**Key words: Proteomics, Precision Medicine, Mass spectrometry, Microsampling, Longitudinal Sampling, Dried blood spots**

### Targeted Proteomics for Precision Medicine

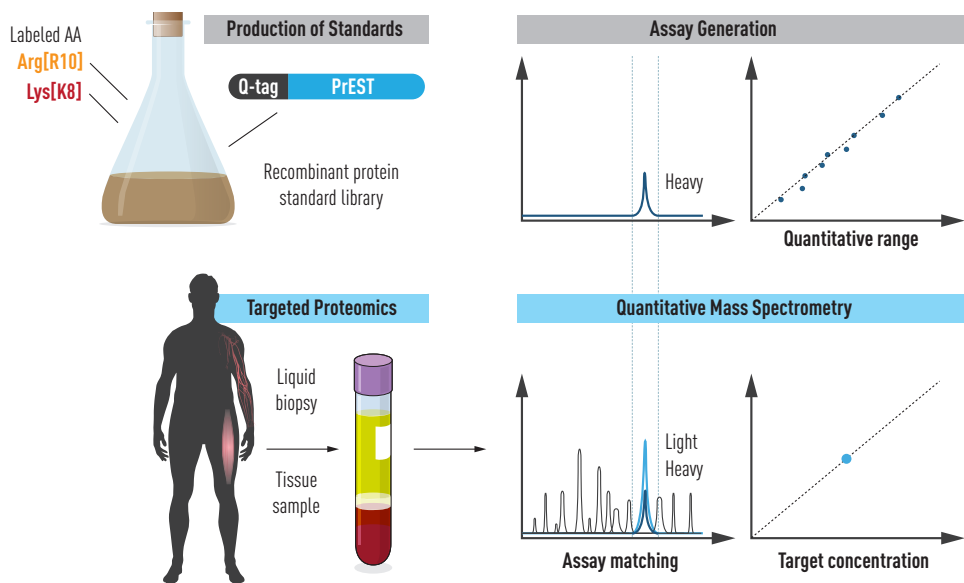
The dramatic improvements have led the field of targeted proteomics in genome sequencing, but methods for robust and accurate analysis of proteins have lacked behind. This delay is due to: 1) prohibitive cost of comprehensive molecular phenotyping and 2) inherent sampling difficulties. Blood is the most lucrative sample source for molecular phenotyping due to its accessibility and systemic nature. However, blood samples necessitate interactions with a phlebotomist, typically in a clinical setting. This requirement simultaneously increases the cost and limits the scale of molecular phenotyping efforts. Therefore, the need for standardization of protein quantification cannot be emphasized enough and the necessity for internal standards to ensure high precision and accuracy is often overlooked. We have established a robust platform for protein quantification that enables recombinant protein fragments to be stored for extended periods of time without compromising the quantitative precision when spiked into blood plasma. Plasma samples to be analyzed are simply added to a dried pellet followed by enzymatic treatment and mass spectrometry analysis.

We have shown that this approach can be used to precisely ( $CV < 10\%$ ) determine the absolute concentrations in human plasma of hundred clinically relevant protein targets, spanning four orders of magnitude, using simultaneous analysis of 292 peptides.

Using this quantitative technology, we can develop a low-cost, scalable molecular phenotyping technology that can be performed independently of a clinic visit. We apply this technology to improve risk prediction and early detection of diseases. by combining genotype data and machine learning. Our core innovations are robust molecular quantification, which can be provide extended phenotype and genotype information to identify new disease subtypes, discover genetic predictors of these subtypes, and predict health trajectories with longitudinal sampling.

### Short Bio

Researcher at KTH working developing next generation proteomics tools for precision medicine applications. My research is driven by mass spectrometry applications and I am working with novel technologies that enable precise and accurate quantification of proteins and metabolites from blood plasma, dried blood spots and self-collected samples at home.





## Sonja Aits

Community: Life Science  
Associate Senior Lecturer, Lund University  
sonja.aits@med.lu.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it,  
Have a research tool or approach and want to find a collaborator with an idea or data to apply it to**

**Research areas: cell and molecular biology, precision medicine and diagnostics, epidemiology and infection biology, AI/MLX, AI/Math**

**Key words: image analysis, text mining, natural language processing, deep learning, systems biology, COVID-19, cancer, neurological diseases, sustainability**

Combining computer vision, natural language processing, systems biology and clinical research

My group combines medical research, bioinformatics and artificial intelligence methods, focusing on these areas:

1. Understanding the regulation of cellular pathways, especially those related to cell death and lysosomes, and how they are affected in diseases and by environmental factors. For this we combine high-content imaging, computer vision, natural language processing and "classical" bioinformatics.
2. Extracting COVID-19-related information from patient journals and scientific literature.
3. Developing artificial intelligence-based natural language processing tools for information extraction from Swedish and English medicine-related texts (e.g. scientific literature, patient journals, social media)
4. Developing artificial intelligence-based image analysis tools for microscopy and histology image analysis
5. Sustainable AI

We work very interdisciplinary and always open to collaborations on both the technical side and life science/medical side, e.g. with groups

- developing new model architectures for computer vision or natural language processing
- developing machine learning for genomic data
- developing graph-based algorithms
- developing models applicable in X-ray and neutron sciences
- developing mature software/web tools
- working with high-content screening and related bioimage analysis



- working with histology scoring
- studying cell death-, autophagy- or lysosome-related processes and diseases (e.g. cancer, stroke, neurodegeneration, lysosomal storage diseases) in humans or other organisms
- studying pollution and its effects on humans and other organisms
- requiring help with text mining of scientific texts or patient journals
- requiring help with image analysis
- requiring help with AI-based methods for sustainability research
- studying AI from a social science or sustainability perspective

Besides this, we are interested in finding partners for AI Lund, an interdisciplinary umbrella organization for research, education and outreach related to AI at Lund University, or with the COMPUTE Research school, which is involved in research and education related AI and non-AI computational tools. We are also very interested in AI-related citizen science and popular science cooperations.

### Short Bio

I lead the "Cell Death, Lysosomes and Artificial Intelligence" group at Lund University and work at the intersection of medicine/life science and artificial intelligence. I am trained originally in biomedicine (PhD from Lund University 2010) and worked as Postdoc and Senior Scientist in lab-based research in Denmark and Australia until 2018, with focus on cell death, autophagy and lysosomes. My group is primarily working computationally and very interdisciplinary consisting of people from computer sciences/engineering, mathematics, bioinformatics, biomedicine and clinical medicine. We are part of the SciLifeLab national COVID-19 program and I also contribute to the development of the Swedish COVID-19 data portal.

I am also one of the coordinators of AI Lund (<https://www.ai.lu.se/>), the umbrella organization for AI-related research, education and outreach at Lund University. Besides this, I am study director of the COMPUTE Research School (<https://www.compute.lu.se/>), where I have developed the "AI in medicine and life sciences" PhD course program. I am also a fellow of the Lund Institute of Advanced Neutron and X-ray Science (<https://www.linxs.se/>).



## Patrick Lambrix

Community: WASP  
Professor, Linköping University  
patrick.lambrix@liu.se



**Looking for open problems/questions to create new collaborations,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: AI/MLX, Software**

**Key words: FAIR data, Semantic access to data, Data integration, Databases,  
Ontologies**

Semantic and integrated access to FAIR data for use in data-driven science

I am interested to cooperate with a DDLS researcher/team (e.g. with joint postdocs or twinning of postdocs) in a project where I can bring my experience in databases and knowledge representation for FAIR data.

Most of my AI research has been performed in knowledge representation and engineering where my aim is to develop methods based on a strong theoretical background that are then implemented in tools that can be used by knowledge engineers and domain experts in different fields. One area in which I have pioneered work is ontology engineering. Ontologies are a key technology for the Semantic Web and for making data FAIR. My group has developed theory and tools for ontology alignment, ontology debugging and ontology completion and we co-organize workshops and competitions in this area.

My work in knowledge engineering has been used in, e.g., European Union projects and the Swedish e-Science Research Center, in areas such as biomedicine, security and materials science. As an example, a recent project is in close cooperation with the international OPTIMADE consortium that gathers the major materials science calculations databases where we provide semantic and integrated access to such databases.

Within the biomedical field we have worked with ontologies in BioPortal, organize the Anatomy track of the OAEI competition, co-developed the Animal/One Health Surveillance Ontology (with Statens Veterinärmedicinska Anstalt), aligned ontologies in toxicology and MeSH for Livsmedelsverket, evaluated approaches for storage, management and integration of molecular interaction data, and worked on semantic access to databases. We have also worked on evaluations of different storage and query facilities for archetype-based electronic health record data.

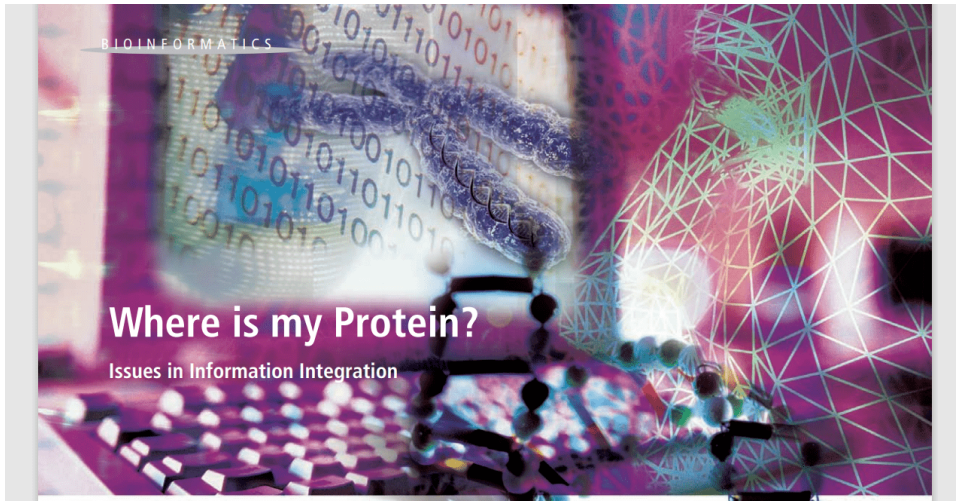
More info at <https://www.ida.liu.se/~patla00/>

Abstracts Machmaking WASP - DDLS 2021

## Short Bio

Patrick Lambrix is professor at the Department of Computer and Information Science of Linköping University. He is the head of the Division for Databases and Information Techniques and the Database and Web Information Systems Group. He is a member of the Swedish e-Science Research Centre and affiliate of WASP.

His main expertise is in knowledge engineering and his current research interests include Semantic Web, Databases, Bioinformatics and Sports analytics. He has organized several conferences and workshops in knowledge engineering, was keynote/invited speaker at several events, received several best paper selections and the Young Scientist Award of the Swedish AI Society.



## Henrik Green

Community: Life Science  
Professor, Linköping University  
henrik.gren@liu.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: cell and molecular biology, precision medicine and diagnostics**

**Key words: Toxicity, adverse drug reactions, chemotherapy, cancer, single cell  
RNASeq, MPS, NPS, Forensics**

Pharmacogenetics, toxicity, personalized chemotherapy by  
prediction of neuropathy and myelosuppression in cancer  
treatment

My group works on predictions models for adverse drug reactions / toxicity using massive parallel sequencing in combination with single cell RNASeq, HiCap and artificial intelligence / machine learning. We especially work with chemotherapy (anti-cancer) induced myelosuppression and neuropathy, so we are an excellent partner for any drug related proposal that also wants to look at toxicity of the drugs/treatment of interest. We would also like to see if we can have an AI researcher working on our data either as joint, twinning or transfer of postdocs.

Treatment with classical chemotherapies induce toxicity and adverse drug reactions (ADRs) leading to unnecessary mortality, suffering and high hospitalization costs. We believe that modern chemotherapy has two cornerstones; predicting response based on tumor mutations and predicting toxicity based on the patients genotype and predisposition. However, predictive models for toxicity is today lacking, especially for myelosuppression and neuropathy, which often are life-threatening and disabling.

The novelty of our research lies in the combination of several approaches;

-Accurate clinical description of toxicities with whole genome data from 212-525 patient co-horts. We use SciLifeLab and NGI to whole-exome and whole-genome sequence cohorts of 342-525 patients treated with combinations of drugs such as gemcitabine, carboplatin and taxanes to investigate the underlying genetics of neuropathy and myelosuppression.

-High resolution chromatin conformation capture identifying interactions between genetic variants, promotor and/or enhancers upon drug exposure. We use HiCap to identify which genetic variants

and genes from our WES and WGS sequencing data that actually are interacting during drug exposure (works excellently).

- Single-cell RNA-seq of human bone marrow to identify pathways involved in tissue specific drug response.

We will combine these data with two different systems biology approaches; protein-protein network modelling (using the most functionally connected genes) and auto-encoders trained on the UK-Biobank exome genotypes of several hundred thousand individuals.

Using these approaches, we hope to establish validated models that can predict patients' risk of toxicity/ADRs in the clinic. We are also open to new and old collaborations.

I also work a lot on new psychoactive substances, NPS, drug such as internetdrugs. Here I have access to case samples from the National Board of Forensic Science, methods to detect NPS drugs, metabolism and mechanism of action. We have access to a huge number of new drugs and are investigating their mechanism of action using CB1, 5HT2a, my-opioid receptor and the monoamnietransporters, SERT, NET and DAT.

## **Short Bio**

Professor at Linköping University, jointly employed with the National Board of Forensic Medicine.

Previously: 2012-2020 Research Strategist / Molecular Biologist in Engineering and Genomics. Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping.

2011-2012 Post-doc at Science for Life Laboratory, Division of Gene technology, Institution for Biotechnology, KTH - Royal Institute of Technology.

2012 Associate Professor in Pharmacogenetics at Linköping University.

2007-2011. Assistant Professor at the Division of Drug Research, Faculty of Health Sciences, Linköping University, Linköping Sweden.

Supervised 7 PhD as main or co-supervisor until dissertation. Supervised 5 post-docs.



## Ola Spjuth

Community: Life Science  
Professor, Uppsala University  
ola.spjuth@farmbio.uu.se



**Have an open project or idea that could benefit from a collaboration**

**Research areas: cell and molecular biology, precision medicine and diagnostics, Autonomous systems**

**Key words: Decision making, drug discovery, lab automation, AI, machine learning**

### Towards autonomous decision making in automated phenotypic drug discovery

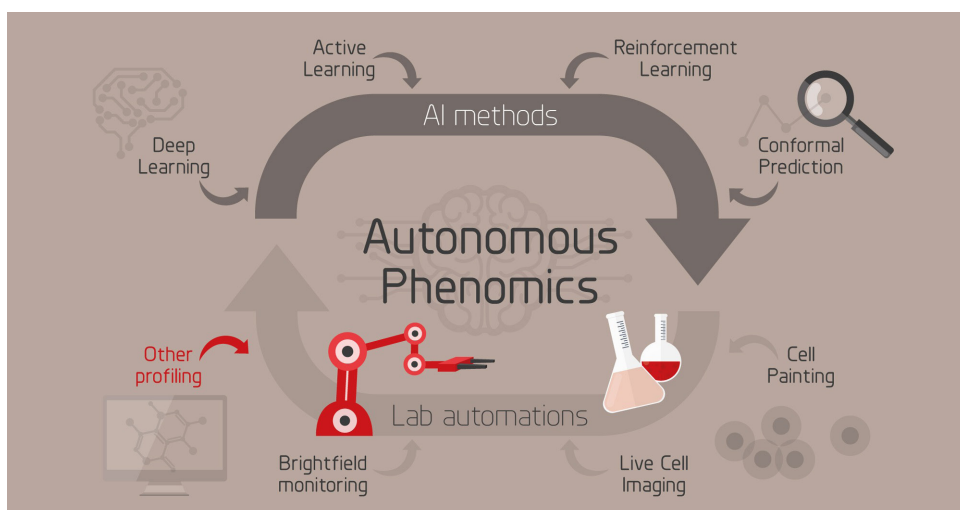
We have set up an open source robotized lab for image-based drug screening and cell profiling. The lab can operate on multiple simultaneous 384-well microplates with hundreds of cells per well, and we use high-content microscopy imaging as the primary readout to capture changes in morphology (Cell Painting protocol) after cells being exposed to treatments with individual or combinations of chemical compounds. This generates very large amounts of images, and we have put much effort into automating QC and analysis pipelines for captured images as well as segmented cells and their morphological profiles. We have already, and are continuing to build up, a large database within different applications in cell biology, environmental toxicology, and precision cancer medicine.

We would now like to take the next step and move towards autonomous discoveries by implementing decision making to design the next round of experiments towards a specific objective; e.g. finding the best combination and concentrations of a set of 3-4 drugs to selectively kill specific cancer cells while not harming normal cells. To this end, we are looking to establish a collaboration with researchers in WASP that can contribute with theoretical/methodological expertise in AI for decision making. From our side we offer a robotized platform to evaluate new AI methods for decision making in a real-world setting with exciting applications in data-driven life science.

Research group website: <https://pharmb.io/>

### Short Bio

Ola Spjuth is Professor in Pharmaceutical Bioinformatics at Uppsala University and AI-coordinator at SciLifeLab Data Center. He leads an interdisciplinary team of 17 people engaged in data-driven cell biology and drug discovery. His research involves robotized and automated high-throughput and high-content molecular and cellular technologies, and how this can be coupled with AI/machine learning for intelligent decision making and analysis to speed up scientific discoveries and enable us to tackle previously unfeasible problems.





Project started!



## Tuuli Lappalainen

Community: Life Science  
Professor, KTH/SciLifeLab/  
New York Genome Center  
tuuli.lappalainen@scilifelab.se



**Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it**

**Research areas: cell and molecular biology, precision medicine and diagnostics**

**Key words: human genetics, genetic variants, transcriptome, proteome, genome, prediction, sequence data, cellular function**

### Novel computational methods for analysis of molecular effects of genetic variants

My research group studies functional genetic variation in the human populations. We are particularly interested in characterizing how genetic variants affect the transcriptome and other molecular phenotypes, which further propagates to changes in cellular functions and physiological changes that underlie diverse human diseases. We analyze these questions both by computational integration of genomic and transcriptomic data sets from cohorts of thousands of individuals and by experimental perturbations of the genome in cellular model systems.

In the context of this joint call, I am interested in new computational collaborations to seek new solutions in the broad area of three fundamental problems that are at the core of my research program: 1) We have extensive data of genetic variants (including variants that affect disease risk) affecting gene expression and transcript structure. However, linking these data to advanced models of protein dosage, structure and interactions is a much less developed area with major potential gains; 2) We are generally interested in novel analysis methods for transcriptome sequencing data, including novel long-read data; 3) We have increasingly complex readouts of cellular functions, including highly multidimensional and multi-modal data, where new representations could uncover novel biology. In addition to these biological problems, we are generally interested in visualization and management of large and complex functional genomics and human genetics data.

### Short Bio

I have recently relocated to Sweden from New York, and I am currently building my lab at the Department of Gene Technology at KTH. I continue run a lab at the New York Genome Center, and I am also the director of SciLifeLab's Genomics Platform and the National Genomics Infrastructure. I am involved in several international consortia and collaborative efforts in human genetics. Being new to the Swedish research community, I am keen to link with potential new collaborators (inside or outside the context of this call). While we use computational approaches, and hard-core methods development and advanced modeling is beyond our area of expertise.



## Maria Lerm

Community: Life Science  
Professor, Linköping University  
maria.lerm@liu.se



**Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it**

**Research areas: cell and molecular biology, precision medicine and diagnostics,  
epidemiology and infection biology**

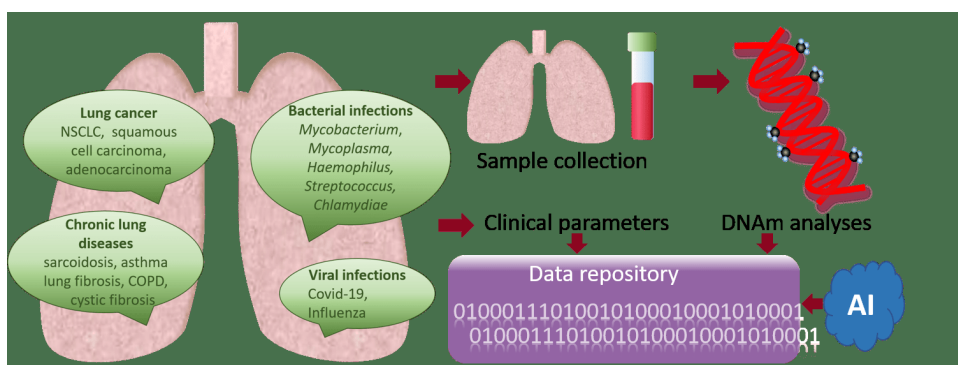
**Key words: DNA methylation, lung medicine, precision medicine, epigenetics, high-dimensional data**

### DNA methylation biosignatures for precision medicine in lung diseases

Epigenetics is a branch of genetics that is gaining increasing interest in research and development. The environment we live in and the diseases we are exposed to affect epigenetic mechanisms in different ways to keep us healthy. Since lifestyle and diseases are reflected in epigenetic markers, e.g. methylation pattern in DNA, these may also be effectively used to diagnose or predict disease or provide information on how a disease condition should be optimally treated. Knowledge in this field has exploded during the last decade, but few findings have so far been translated to the clinical setting. We have accumulated DNA methylation data over the years, mainly from clinical studies on infectious diseases including Covid-19 and tuberculosis. We have trained bioinformaticians within the group, who are including machine learning in their analyses, but our next step will include high-dimensional data, clinical data, lab data along with DNA methylome analyses, which will require more powerful analysis algorithms. The aim of our studies is to provide health care with diagnostic precision tools based on DNA methylation panels. Also, we expect to find new disease associations and improve our understanding of disease mechanisms. Our research is focused around respiratory medicine.

#### Short Bio

I had my basic training in Biology at Linköping University and in 1996 I moved to Germany for PhD studies. In 2002, I was back in Linköping for post doc studies at the Faculty of Medicine and Health Sciences, where I am professor today. My research focus has always been host-pathogen interaction and our work comprise basic investigations of cellular and molecular events at the host-pathogen interface, as well as clinical studies in Sweden and in Peru. In all areas, we have a special focus on epigenetics. My group consists of three biologists and two bioinformaticians with PhD grades, three PhD students and several undergraduate students.





## Tanja Slotte

Community: Life Science

Associate Professor in Ecological Genomics

[tanja.slotte@su.se](mailto:tanja.slotte@su.se)

21

**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it,  
Looking to collaborate with experts from another field to develop new methods**

**Research areas: evolution and biodiversity**

**Key words: Machine learning, Population genetics, Genomic data, Simulations, Adaptation, Inbreeding**

### Machine learning for population genomics

In population genetics, machine learning methods have recently been shown to outperform many classical analysis approaches when it comes to identifying genomic regions of adaptive significance. However, as currently implemented, these methods are not fully applicable to data from populations that are partially inbred or that have undergone strong changes in population size. Developing methods that perform better under these circumstances is important and timely as climate change and increasing habitat fragmentation is predicted to lead to reductions of population sizes and increased inbreeding levels in many species. To be able to robustly identify genomic regions of adaptive significance in such populations could be important for conservation of biodiversity. As part of research projects funded by the ERC, the Swedish Research Council and the SciLifeLab National Biodiversity programme, my group has already generated population genomic data sets useful for testing new machine learning methods. We have experience in computational analyses of large-scale genomic data and large-scale population genetic simulations, and we have started running existing machine learning methods. Now we are interested in taking this research line further to develop new methods. I am therefore looking for a collaborator with expertise in machine learning for a WASP-DDLS application within the field of Evolution and Biodiversity.

### Short Bio

I am an Associate Professor in Ecological Genomics at Stockholm University. I am a population geneticist by training and in my work I use genomic data to test evolutionary hypotheses. I received my PhD from Uppsala University in 2007, followed by a postdoc at University of Toronto. After that I moved back to Sweden to start my own group at Uppsala University. In 2014 I took up a position as a SciLifeLab Fellow affiliated with Stockholm University.

For more information, see <https://tanjaslottelab.se/>



## Simon Elsässer

Community: Life Science  
Dr., KI  
simon.elsasser@scilifelab.se



**Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: cell and molecular biology, precision medicine and diagnostics**

**Key words: epigenomics, high-throughput, drug screening**

### Highly multiplexed quantitative ChIP for epigenetic drug profiling and toxicity assessment

We are developing a highly multiplexed quantitative ChIP platform for profiling how the epigenome responds to drugs. We want to know if drug action perturbs epigenetic regulation in the short term or even long term, and if a cell-based high-throughput assay (e.g. primary cell lines, cancer cell lines, 3D spheroid/organoid model) can predict drug efficacy and toxicity in the human body. We believe that comprehensive epigenome profiling will be part of the future drug development pipelines to better score and select lead compounds for animal and human trials.

**OFFERING:** 1) various existing medium-scale quantitative ChIP-Seq datasets (tens-100s individual genome-wide datasets). 2) Developing a method to measure up to 24 epigenome markers (e.g. histone modifications) versus 96 drug candidates

**SEEKING:** collaborators for developing computational framework to analyze large-scale epigenome datasets in response to drug action and to generate predictive algorithms based on profiling existing drug candidates with known toxicity data from successful or failed clinical trials.

### Short Bio

SciLifeLab Fellow at Campus Solna

## Kevin Smith

Community: WASP, Life Science  
Associate professor, KTH, SciLifeLab  
ksmith@kth.se



### **Looking for open problems/questions to create new collaborations**

#### **Research areas: AI/MLX**

**Key words:** machine learning, AI, computer vision, bioimage analysis, bioimage informatics

### Computer vision and bioimage analysis

I work at the intersection of machine learning, computer vision, and life science. I am interested in how machine intelligence can help solve vital questions in biology and medicine, in particular through the quantification and understanding of patterns in images.

If you have a project what could benefit from image analysis with AI, please feel free to contact me.

#### **Short Bio**

I am an associate professor in computer vision and biomedical image analysis with the [KTH Royal Institute of Technology](#) and the [Science for Life Laboratory](#) Solna campus.

I am interested how machine intelligence can help solve pressing questions in biology and medicine, in particular through the understanding and quantification of patterns in images. Recently, an explosion of digitalized data, groundbreaking advances in the life sciences, and the development of cutting-edge algorithms capable of understanding biomedical data, especially from images, have generated terrific excitement around AI and healthcare. My research aims at developing intelligent systems that can precisely diagnose conditions, reduce patient risk, choose effective treatments, and further our understanding of biological systems.





## Björn Wallner

Community: Life Science

Professor

[bjorn.wallner@liu.se](mailto:bjorn.wallner@liu.se)

24

**Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it**

**Research areas: cell and molecular biology**

**Key words: protein structure, protein-protein interactions, deep reinforcement learning, graph networks, rotation invariant learning**

### Using AI to reason over protein-protein interactions

Proteins are molecular machines that are responsible for most processes in life. As with any machine the components and the structure of the machine is what makes it operate. For proteins the building blocks are 20 amino acids with different properties that when placed in the right order (the protein sequence, 20-1000 amino acids) folds into a unique 3D structure capable of performing very specific functions, like binding the oxygen molecules you used breathed in. The protein sequence is in most cases easy to obtain from experiment but the mapping from the 1D sequence to the 3D structure much harder. Demonstrated by the fact that there are roughly 214M known protein sequences but only 0.175M known protein 3D structures. Adding to the complexity is that proteins are not working in isolation, but they are social characters that perform their function by interacting with other proteins, macromolecules (RNA or DNA) or small molecules (metals, drugs) as well as the solvent (water or lipid membranes) often in a highly dynamic manner where flexible regions of the proteins are moving in the response to its surroundings.

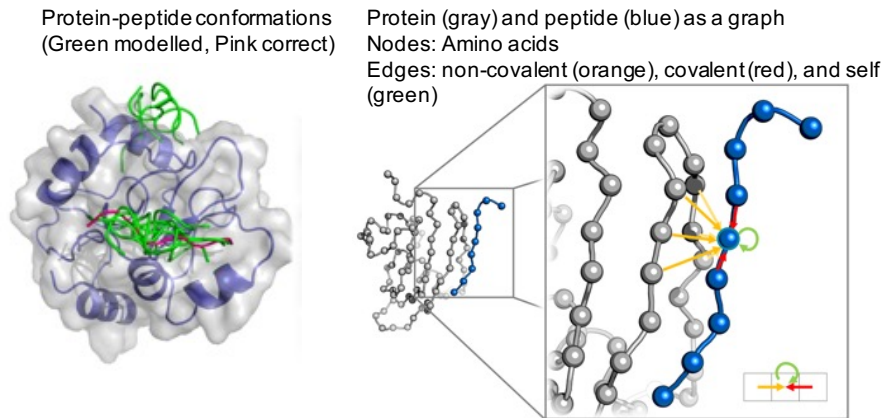
We are interested protein flexibility/disorder and how it is in regulating life at the molecular level. Last year DeepMind claimed to have solve the protein folding problem using the deep learning method AlphaFold2, showing that it is actually possible to apply advanced machine learning methods to this problem domain. So far AlphaFold2 is not publically available so it is yet to be seen actually how useful it will be for the scientific community. Still, it would also not work on protein flexibility and interactions without significantly changes to the algorithm.

We want to use AI method that are able to reason and learn traits of protein-protein interactions and that are able to generate protein-protein conformations. In the past we have used graph nets to represent the structure of a protein, which is a natural representation since you can impose the contact network in the protein as a graph, and we are using an external modelling software to generate potential conformations for the network to judge. We are now using deep reinforcement learning to optimize the ability of the network graph to rank conformations of a sampling trajectory.



## Short Bio

Structural bioinformatician with 20 years of experience of machine learning in protein structure prediction field. He works close with experimental groups and use computational methods and developed ML-methods to integrate different types of data with predictions and simulations. He is involved in developing the Rosetta software suite for modeling macromolecules and has developed a number of methods that are world leading. The ML-based predictor of local coordinate error of protein structure models has been top-ranked in the community-wide benchmarks CASP11 (2014) and it has been used to solve the crystallographic phase problem to enable structure determination of more proteins.



## Andreas Kerren

Community: WASP  
Professor, Linköping University  
andreas.kerren@liu.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: AI/MLX, Software, Non applicable**

**Key words: visual analytics, information visualization, biological networks,  
exploration, data analysis, text analytics, network analytics, visualization**

### Visual Network and Text Analytics for the Life Sciences

The **visualization and analysis of network structures** is a very prominent research topic because networks play a decisive role in the life sciences, such as the visualization of large and complex biological networks. The overall problem of scalability is not only limited to the number of nodes or edges. Due to the sheer network size and complexity, their growth and continuous change, as well as their compilation from databases on demand, life scientists very often request novel and effective network visualization, interaction and exploration techniques. Moreover in systems biology, there is a hierarchy of biological networks that shape sets of heterogeneous networks which form layers. This adds even more problems to the task of network analysis. All those challenges, ranging from the visual analysis of multilayer networks to multivariate (and eventually temporal) networks, are equally important for the visualization community and for the life sciences.

Similarly challenging questions can be found at the intersection of NLP (i.e., text analytics) and the life sciences. As Larry Hunter already mentioned in his keynote at BioVis 15: "... the average human gene now has more than 2,000 publications about it. In order to benefit fully from existing knowledge, researchers must assimilate extensive information from many domains, often well outside their own areas of expertise. The amount, diversity, and complexity of all this potentially relevant information puts full exploitation of it beyond the time and attention capacity of even dedicated researchers. The information visualization community needs to develop new user-specific information visualization approaches that focus biomedical researchers' attention on the material most interesting to them." This statement is still valid today. Bringing visualization into the play (and as such the human back into the analytical loop) will make it possible for life scientists to **make sense of large and dynamic text data** and allows for exploration, control and final evaluation of the analysis processes and results.

Both areas are also strongly related to machine learning, explainable ML/AI, and multidimensional data analysis. We are very much interested in close collaborations with the life science community

to develop together approaches and techniques for addressing the above mentioned (or similar) challenges. We are also open for any other proposals from the life sciences where there are challenging visual analytics problems. Some additional information can be found here:

[A Visual Guide to BioVis Techniques](#)

[Network Visualization for Integrative Bioinformatics](#)

[A Visual Survey of Text Visualization Techniques](#)

### **Short Bio**

Prof. Dr. Andreas Kerren received his PhD degree in Computer Science from Saarland University, Saarbrücken, Germany. In 2008, he achieved his habilitation (docent competence) from Växjö University, Sweden. Dr. Kerren is currently a Full Professor of Information Visualization, Linköping University (LiU) and Linnaeus University (LNU), Sweden. He holds the Chair of Information Visualization at LiU and is head of the research group Information and Software Visualization at LNU. His main research interests include several areas of information visualization and visual analytics, especially visual network analytics, text visualization, and the use of visual analytics for explainable AI.



## Maria Sandsten

Community: WASP  
Professor, Lund University,  
Centre for Mathematical Sciences  
maria.sandsten@matstat.lu.se



**Looking for open problems/questions to create new collaborations,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: AI/Math**

**Key words: Time-frequency representations, Spectrum analysis, Transient signals,  
Oscillatory structures, high precision, robust, source localization**

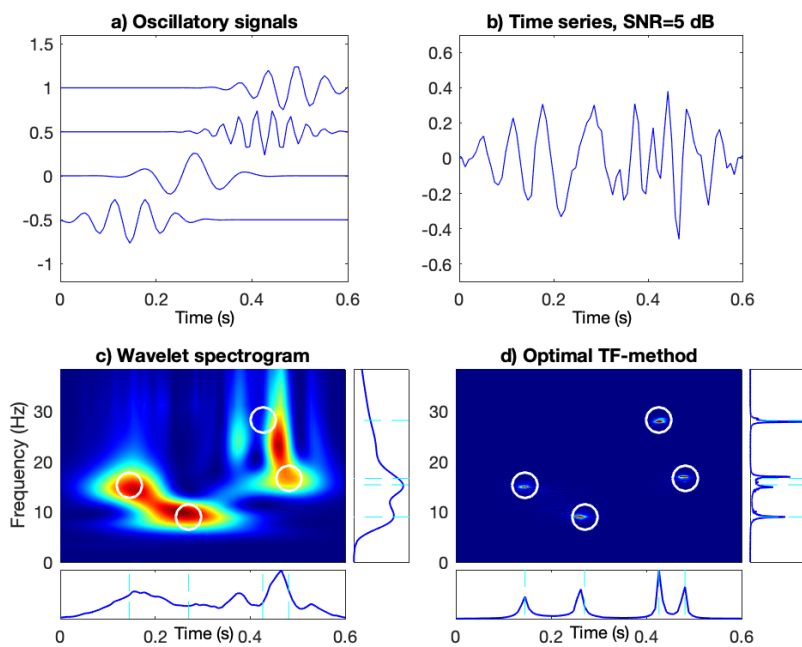
### Classification and characterization of non-stationary time-series using optimal time-frequency analysis and machine learning

Non-stationary oscillatory signals are the key structures of measured time-series in many application areas. Expanding such signals into time-frequency representations/images is often necessary to extract relevant information, for example to extract features to classify the signals. We will serve research with optimal and tailored time-frequency methods, which have the ability to resolve important and hidden information of such signals. We have expertise in the complete toolbox of time-frequency analysis, as well as statistical methods for investigating properties of multi-dimensional signals, such as causality, connectivity and source reconstruction.

Recent research areas where we have applied our methodology is machine learning-based classification of oscillatory structures in electroencephalogram data. The characterization of the mental states of the brain is challenging due to the high amount of noise. In this context phase synchrony and cross-frequency coupling measures are defined as important features. We also have collaboration in the area of source separation using time-frequency analysis methods and deep learning networks for acoustic signal separation. Further we have designed methods for optimized resolution of close components in ultrasound radio frequency data. We also collaborate on underwater acoustic signal characterization, with the aim to discover how dolphins use their ultrasonic echolocation beam and classification of bird song syllables.

#### Short Bio

Professor in Statistical Signal Processing.



# Stefan Bauer

Community: WASP  
Dr. KTH  
stefan.a.bauer@gmail.com



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: AI/MLX, Autonomous systems**

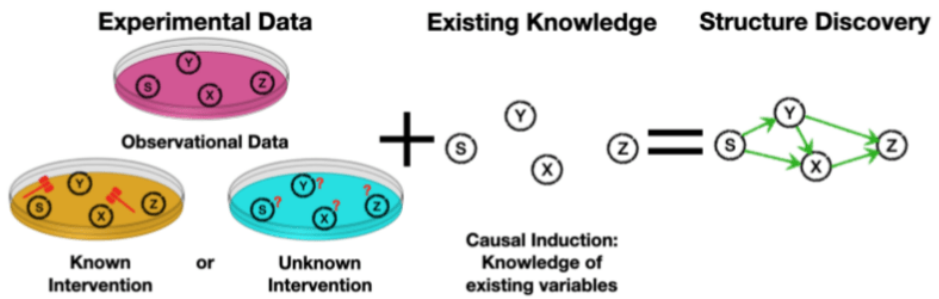
**Key words: Causal Inference, Representation Learning, Healthcare**

## From Prediction to Causation

Many questions in everyday life as well as in research are causal in nature: How would the climate have changed had we reduced our carbon emissions in the 80s? Will my headache go away if I take an aspirin? Inherently, such questions need to specify the causal variables relevant to the question. A central problem for AI and key application areas like health care or robotics is thus the discovery of high-level causal variables from low-level observations like pixel values. While deep neural networks have achieved outstanding success in learning powerful representations for prediction, they fail to explain the effect of interventions. This is reflected in a limited ability to transfer and generalize even between related tasks. As a way forward we propose to learn causal representations from data and our recent efforts combine interventions and causal structure with deep learning based approaches. Using and developing tools of causality, deep learning and real robotic systems, my research group focuses on the longstanding goal of artificial intelligence to design machines that can extrapolate experience across environments and tasks.

## Short Bio

Stefan Bauer is an Assistant Professor at KTH Stockholm, and a CIFAR Azrieli Global Scholar. Previously he was a research group leader at MPI for Intelligent Systems. He obtained his PhD in Computer Science from ETH Zurich and was awarded with the ETH medal for an outstanding doctoral thesis. Before that, he graduated with a BSc and MSc in Mathematics from ETH Zurich and a BSc in Economics and Finance from the University of London (LSE). During his studies, he held scholarships from the Swiss and German National Merit Foundation. In 2019, he won the best paper award at the International Conference of Machine Learning (ICML) and in 2020, he was the lead organizer of the [real-robot-challenge.com](https://real-robot-challenge.com), a robotics challenge in the cloud.





## Linn Fagerberg

Community: Life Science  
Group leader, KTH, SciLifeLab  
[linn.fagerberg@scilifelab.se](mailto:linn.fagerberg@scilifelab.se)



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it**

**Research areas: precision medicine and diagnostics**

**Key words: Bioinformatics, Data integration, Precision medicine, Multi-Omics, Proteomics, Transcriptomics**

### Integration of omics data in life science

The main research focus of my group is integrative multi-omics where large-scale biological data is analyzed using advanced bioinformatic methods. The integrative omics research is both performed within the precision medicine field as well as within the framework of the Human Protein Atlas.

The Human Protein Atlas program allows for a genome-wide exploration of the protein-coding genes expressed across human immune cell populations and all major tissues and organs. My group is responsible for the quantitative transcriptomics-based classification and gene-based clustering of the human proteome, as well the integrative approaches mainly involving proteomics and transcriptomic data. The expression-based analyses have been performed on the global tissue level as well as in single cell types and we have also characterized the human immune cells and the different regions of the human brain.

An important aspect of precision medicine is to probe and define the differences in molecular profiles among healthy and diseased individuals. The Swedish SCAPIS SciLifeLab Wellness Profiling Program (S3WP) is based on the Swedish CArdioPulmonary bioImage Study (SCAPIS) and aims to study the longitudinal effects of lifestyle variation based on personalized omics profiles. The participants of the S3WP program are profiled based on a combination of classical clinical chemistry, advanced medical imaging and extensive omics profiling, including the analysis of the genome, transcriptome, plasma proteome, plasma metabolome, blood cell composition (immune cytome), auto-antibody reactivity profiles and gut microbiota. We started out by following 101 healthy individuals longitudinally for two years and are continuing with disease cohorts including type 2 diabetes and cardiovascular disease. We are also profiling the blood proteome in premature children to obtain a deeper understanding of the preterm infant.

We now intend to expand the omics-based profiling across a wide range of diseases based on well-characterized cohorts, with the aim to characterize molecular fingerprints and allow stratification of patients to pave the way for future precision medicine efforts.

Abstracts Machmaking WASP - DDLS 2021



Selected publications:

Uhlen et al (2019) "A genome-wide transcriptomic analysis of protein-coding genes in human blood cells" *Science* 366 (6472): eann9198

Zhong et al (2020) "Whole-genome sequence association analysis of blood proteins in a longitudinal wellness cohort." *Genome Med.* 2020 Jun 23;12(1):53.

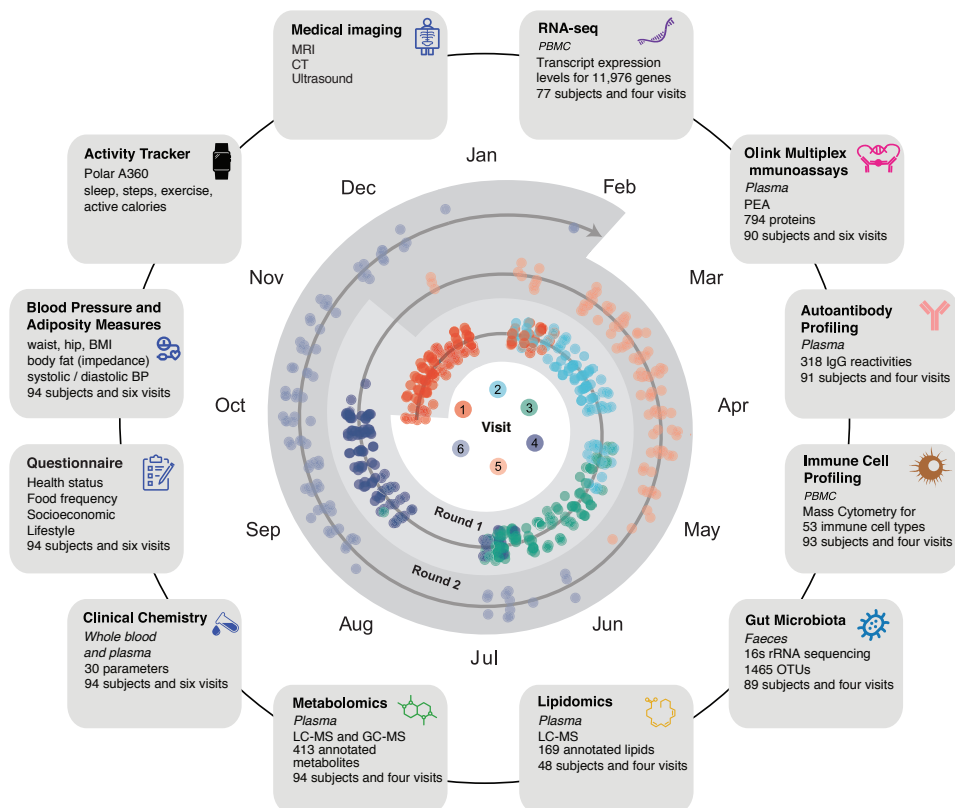
Tebani et al (2020) "Integration of molecular profiles in a longitudinal wellness profiling cohort." *Nat Commun.* 2020 Sep 8;11(1):4487.

Zhong et al (2021) "Dramatic changes in blood protein levels during the first week of life in extremely preterm infants." *Pediatr Res.* 2021 Feb;89(3):604-612

Zhong et al (2021) "Next generation plasma proteome profiling to monitor health and disease." *Nat Commun.* 2021 May 3;12(1):2493

## Short Bio

I am a researcher at KTH with a background in bioinformatics engineering and a key interest in data visualization. My main focus has been on data analysis within the Human Protein Atlas program. In the past few years, my research has moved into the field of precision medicine to allow for profiling of human health and disease based on multi-level integrative omics analyses.





## Pontus Nordenfelt

Community: Life Science  
Associate Professor, Lund University  
name.name@scilifelab.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: ccell and molecular biology, precision medicine and diagnostics,  
epidemiology and infection biology, AI/MLX, Autonomous systems, Software.**

**Key words: automated and autonomous high-resolution microscopy, feedback  
microscopy, correlative microscopy, advanced live imaging**

### Data-driven microscopy

High-resolution microscopy has issues with operator bias, reproducibility, and sample heterogeneity that are limiting the quantitative aspect of the technique. We have developed a modular framework, data-driven microscopy (DDM), that allows for high-resolution imaging of targeted phenotypes in a data-driven manner. The general framework is applicable to any motorized microscope, and all steps can be fully automated, including phenotype targeting. DDM is a useful approach for eliminating human bias in microscopy, increase assay reproducibility, and place single phenotypes in the context of the whole sample population when interpreting high resolution image data.

Current applications that we have implemented in DDM include correlative microscopy where live imaging-based population phenotypes are linked to single-cell high-resolution data, using high-resolution single-cell data to improve low-resolution population data, and targeted imaging across imaging modalities based on population-wide data. Another key aspect of DDM is the server-based architecture with standardized metadata processing, which allows for easy access to large datasets for data mining, AI training data, and online machine learning for feedback microscopy.

We are in the process of releasing the DDM framework to the community and are looking for collaborators that can expand the usage of the methodology, particularly with respect to AI-driven applications.

Seeking: collaborators to enhance the AI aspects of our DDM methodology

Offering: access to large cell biology and infection microscopy data sets, DDM methodology

## **Short Bio**

Pontus Nordenfelt leads a group focused on infection immunology, cancer cell biology, and systems microscopy. He is an engineer by training (LTH), received his PhD at the Medical Faculty in Lund, and several years postdoc at Harvard Medical School (Timothy Springer). Postdoc training as a Fellow at the Harvard Image and Data Analysis Core and alumni of the renowned Woods Hole Physiology Course in advanced quantitative microscopy. He is a member of the Sweden Young Academy (SUA).



## Erik Svensson

Community: Life Science  
Professor, Department of Biology,  
Lund University  
erik.svensson@biol.lu.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it**

**Research areas: evolution and biodiversity**

**Key words: evolutionary ecology, insects, natural selection, Odonata, phenomics, polymorphisms, sexual selection**

### Phenomics, micro- and macroevolution in an old insect order

Research in my laboratory is focused on understanding how microevolutionary processes translate in to macroevolutionary patterns and shape diversification. We are especially interested in how complex phenotypes evolve, selection on character combinations (correlational selection) and the maintenance of phenotypic and genetic variation within natural populations. We are evolutionary ecologists and population biologists who are interested in how natural and sexual selection interact and how frequency-dependent selection and sexual conflict maintains genetic polymorphisms in natural populations.

To achieve our research goals, we combine field experiments, mathematical modelling, quantitative genetics, phylogenetic comparative methods, genomics and transcriptomics. We have a strong theoretical basis in our research laboratory, although we are primarily empiricists who aim to test theoretical predictions in the field and in the lab.

Our main study organisms is an old insect order called Odonata (dragonflies and damselflies). This is an insect order consisting of c. a. 6400 extant species globally, and with a rich fossil record. Odonata is characterized by pronounced intra- and interspecific variation in size, shape, colouration, ecology and behavior. They also have an unusually rich fossil record with preserved wings from about 300 million years BP. The odonate wing is a complex character, consisting of many different veins of unknown function. Ongoing work in my group aims to quantify intra- and interspecific variation in wing morphology and positions of these veins, with the hope that we can better understand evolvability and how microevolutionary processes within species translates in to macroevolutionary diversification between species.

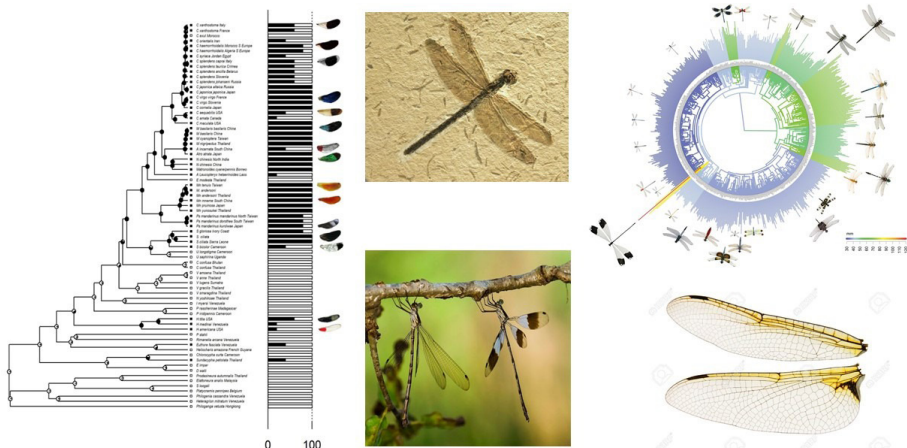
Another strong focus of research is a long-term population biological study of the Common Bluetail damselfly (*Ischnura elegans*) across a series of populations in southern Sweden. This is a polymorphic species with three heritable female colour morphs, governed by one major locus with three alleles in a

dominance hierarchy. This polymorphism is maintained by frequency-dependent sexual conflict due to male mating harassment of common morphs. We have followed populations of this species for more than 20 years and have a unique long-term database consisting of > 60 000 individuals of known sex, age and colour morph, and with fitness data that we can utilize for time-series analyses to infer evolutionary processes. We are also actively working on characterizing this genetic polymorphism at the genomic level, using Next-generation sequencing (NGS) techniques and plan to use CRISPR-CAS in the near future.

We are open to collaboration with scientists in other fields who can complement our core expertise.

## Short Bio

I am an evolutionary biologist, interested in the interface between ecology and evolution. My research has a strong theoretical foundation, but I am also an empiricist interested in evolutionary processes in natural populations. Research in my laboratory combines field and laboratory experiments, evolutionary theory and mathematical modelling with quantitative genetics, molecular and genomic methods. We are especially interested in the link between micro- and macroevolution, evolvability and genetic constraints, causes of long-term stasis and the maintenance of genetic variation through balancing selection caused by sexual conflict in natural populations.



## Maria Fällman

Community: Life Science  
Professor, Umeå University  
maria.fallman@umu.se



**Have a dataset and want to see if there is more that can be done with it**

**Research areas: epidemiology and infection biology**

**Key words: Antimicrobials, bacterial stress responses, bacterial gene expression, chronic infections, antibiotics**

### New strategies to combat bacterial infections

The project addresses basic and applied questions connected to the global threat of increased antibiotic resistance among pathogenic bacteria. Chronic infections and associated antibiotic overuse are well-recognized clinical problems that contribute to this development. Existing antibiotics act on a limited number of bacterial pathways and expanding the set of druggable bacterial components is urgently needed.

One attractive group of targets for new treatment regimens are gene products essential for pathogens to resist at its infection site, where their inactivation would make bacteria escape from its niche. Here, gene products involved in bacterial stress responses altering bacterial gene expression to meet external challenges that commonly occur in host tissue are of particular interest. However, bacterial stress responses are complex and differ depending on infection niches and bacterial subgroups, and their expression patterns and regulations at the in vivo infection site is not fully understood. This high level of complexity is a major obstacle that has hindered a holistic understanding of these mechanisms, and a multidisciplinary approach is clearly desired.

The Fallman lab has experience from studies of bacterial virulence mechanisms and stress responses, and they have recently succeeded to obtain gene expression data from bacteria during infection of host tissue indeed showing that genes connected to environmental adaption dominates during chronic stages of infection. They have also built the PATHOgenex interactive and experimentally validated database with gene expression data from 32 human bacterial pathogens at 11 infection related stress conditions (<http://www.pathogenex.org>; Avican et al, 2021, Nat Comm 12:3282).

We now aim to explore the data in PATHOgenex utilizing new state-of-the-art methods in systems biology including network interference algorithms to identify key components in bacteria that allows adaption to specific conditions. In a next step we will combine the in vitro data in PATHOgenex with global in vivo transcriptomes obtained from bacteria infecting human tissue to identify gene expression patterns involving gene products of critical importance for bacteria to maintain infection in human tissue. The data sets will ultimately be analysed to identify infection-specific gene expression patterns with the goal to identify gene products that can be served as targets for novel antimicrobials.

## Short Bio

Background in Molecular Biology, PhD from 1992 Linköping University, now at Dept of Molecular Biology at Umea University.

Experience from studies of bacterial virulence mechanisms and stress responses. Has together with Dr Kemal Avican, obtained gene expression data from bacteria during infection of host tissue indeed showing that genes connected to environmental adaption dominates during chronic stages of infection (Avican et al., 2015, PlosPathogens). We have also built the PATHOgenex interactive and experimentally validated database with gene expression data from 32 human bacterial pathogens at 11 infection related stress conditions (<http://www.pathogenex.org>; Avican et al, 2021, Nat Comm 12:3282).



## Kristoffer Sahlin

Community: Life Science  
Assistant Professor, Stockholm University  
ksahlin@math.su.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: cell and molecular biology, Software**

**Key words: Sequence similarity searches, algorithms, parametrization**

### Parameter guidance for efficient sequence similarity searches

Our group focuses on algorithm design with applications to biological sequence analysis. This includes identifying sequence similarity, clustering of biological sequences, and sequence error correction. For example, we have previously designed methods to better detect transcripts present in patients with Alzheimer's disease and to find mutated genes in complicated gene clusters with many similar gene copies.

Recently, we have developed a new type of sequence similarity search method [1] that has been demonstrated to be more efficient than state-of-the-art for a popular type of sequence similarity search, namely finding the correct place of a short DNA sequence on a genome [2]. While the general method has been outlined in [1,2], it is still at a proof-of-concept level. Many questions regarding the best parameters and protocol design remain to be explored for different applications with different sequence lengths, biological divergence levels, etc.

We hypothesize that such parameters could be learned using machine learning. We also hypothesize that our sequence matching technique can be applied to general texts and may be of use in other branches of Computer Science and Machine learning such as text mining and Natural language processing.

The potential outcome is a better service for DNA, RNA and protein sequence similarity searches. A well-designed automated parametrization service can improve over popular tools with tens of thousands of citations such as BLAST or DIAMOND, which are used by a large biological and bioinformatic research community.

[1] Kristoffer Sahlin. Strobemers: an alternative to k-mers for sequence comparison. bioRxiv, 2021. doi: <https://doi.org/10.1101/2021.01.28.428549>

[2] Kristoffer Sahlin. Faster short-read mapping with strobemer seeds in syncmer space. bioRxiv, 2021. doi: <https://doi.org/10.1101/2021.06.18.449070>



## **Short Bio**

### **Group research interests**

We use statistical modeling and algorithms to develop methods for analyzing large biological datasets. Particularly, we develop scalable algorithms for high-throughput genomic and transcriptomic sequencing data to study problems related to genome assembly, structural variation detection, and transcriptome analysis. We emphasize applicability of the methods and models to relevant biological and biomedical questions.

### **PI Bio**

I am an Assistant Professor at Stockholm University (the Department of Mathematics) and SciLifeLab Fellow at the national center for molecular biosciences, Science for Life Laboratory. I obtained my PhD in Computer science from KTH Royal institute of Technology and I have worked as a Postdoctoral researcher at Pennsylvania State University and at University of Helsinki. Before that I obtained a bachelor's degree in mathematics and a master's degree in mathematical statistics from Stockholm University, Sweden.

<https://sahlingroup.github.io/>



## Anne-Marie Fors Conolly

Community: Life Science  
Group leader, Umeå University  
anne-marie.fors.conolly@umu.se



**Have an open project or idea that could benefit from a collaboration, Have a dataset and want to see if there is more that can be done with it**

**Research areas: cell and molecular biology, precision medicine and diagnostics, epidemiology and infection biology**

**Key words: Registry data, epidemiology, infectious diseases, complications, risk factors**

### National registry data for infectious diseases - acute and long-term complications

My research group studies the risk factors of and acute and long-term complications following viral infectious diseases such as COVID-19, Hemorrhagic fever with renal syndrome and Influenza.

I have compiled nationwide registry data for all COVID-19 patients and control individuals. Currently I have expanded this project to include the whole Swedish population as controls. Data is obtained from Statistics Sweden, National Board of Health and Welfare, Swedish Intensive Care Registry, Social Insurance Agency Sweden and SKR. There is historical registry data for all hospitalizations for all causes since 1987 for COVID-19 patients, and also all contact with outpatient care and prescribed pharmaceuticals. This data is the foundation for a multitude of projects that I lead and has currently resulted in a publication in The Lancet (Katsoularis et al; 2021) and BMJ Global Health (Fonseca-Rodriguez 2021). These two publications show the breadth of the possibilities of projects using the data as base. The first article focuses on the risk of acute cardiovascular events following COVID-19; and the second article analyzes the spatial clustering of COVID-19 related hospitalization and deaths.

Using this complex data set, I would like to find a collaboration partner who has experience with population studies. I am in the process of obtaining data of familial relations for the Swedish population (thereby also COVID-19 patients) as a proxy for inherent genetic factors as a risk factor for severe disease outcome. Other collaboration interests would be to use machine-based learning, AI or other methods to study large-scale complex data. Since I am also in the process of obtaining data from Social Insurance Agency Sweden (Försäkringskassan) with text, a research group with expertise in retrieving sections of text would be of interest as well.

## Short Bio

I have a Bachelor of Biology and Master's of Science from Copenhagen University. My PhD focused on molecular pathogenic aspects of a viral hemorrhagic fever at Karolinska Institutet. I moved to Umeå to do a postdoc in infectious diseases, and in parallel with research studied medicine to become an authorized clinician in 2020. Currently I am specializing in Clinical Microbiology, however I have mostly been focusing on research, and I am a MIMS clinical research fellow.

My research group consists of a pre-clinical postdoc, two clinical PhD students (main and co-supervisor), a statistician (part-time) and an epidemiologist. The expertise of my group covers cell biology, clinical infectious diseases and statistics and epidemiology.



## Paul Hudson

Community: Life Science  
Associate Professor in Metabolic  
Engineering, KTH  
paul.hudson@scilifelab.se



**Have an open project or idea that could benefit from a collaboration,  
Have a dataset and want to see if there is more that can be done with it,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: cell and molecular biology, evolution and biodiversity**

**Key words: biochemistry, carbon fixation, protein evolution, autocatalytic cycles**

### New technologies to improve biological carbon fixation

Research in my group is focused on the Calvin cycle, the primary metabolic pathway on Earth for fixing carbon-dioxide. In broad terms, we are interested in using machine learning to enhance the function of the Calvin cycle in both in vivo (in microalgae and plants) and in vitro (cell-free systems). Enhanced biological carbon fixation platforms could find use in faster growing crops, or industrial conversion of carbon dioxide into chemicals.

Currently, my group is developing a method to mutate Calvin cycle enzymes in microalgae and screen variants for altered carbon fixation using deep sequencing (see ref 1 for an early version of this approach). What we are looking for in this call is computational expertise in compiling and analyzing the resulting sets of genotype-phenotype associations. The initial analysis is expected to guide targeted engineering of Calvin cycle enzymes, in a so-called directed evolution cycle. The novelty of the approach is that, in contrast to typical laboratory evolution, we will track the fitness of thousands of mutants, which could be used to construct phylogenetic trees during the course of evolution and to identify epistatic relationships for amino acids within key enzymes (e.g. Rubisco) that were previously unknown. This could lead to a breakthrough in engineering the kinetic properties of these enzymes. Expertise in machine-learning would be critical for this approach.

In a complementary research direction, we envision using the experimental data to refine structural kinetic models of the Calvin cycle (see 2 for an example of such models). These models will allow us to identify metabolite and kinetic boundaries within Calvin cycle that allow stable operation. Such models could be critical for establishing carbon fixation outside of the cell.

The project could involve twinning of postdocs with a WASP group, or joint postdoc between us and a WASP group.

Related research:

1. Pooled CRISPRi screening of the cyanobacterium *Synechocystis* sp PCC 6803 for enhanced industrial phenotypes. *Nat Commun.* 2020 Apr 3;11(1):1666. doi: 10.1038/s41467-020-15491-7.
2. Kinetic modeling of the Calvin cycle identifies flux control and stable metabolomes in *Synechocystis* carbon fixation. *J Exp Bot.* 2019 Feb 5;70(3):973-983. doi: 10.1093/jxb/ery382.

### Short Bio

I am an Associate professor at KTH/SciLifeLab (Ph.D. 2009 Chemical Engineering). My research is in experimental metabolic engineering, mostly of bacteria that fix carbon dioxide, for bioenergy or industrial biotechnology applications. We have expertise in synthetic biology (e.g CRISPR/Cas mutagenesis), systems biology (e.g. deep sequencing of mutant populations), and basic metabolic modeling. We are currently looking to collaborate with groups with expertise in machine learning and computational modeling and open to new perspectives/methods in analysis of large datasets. See more at <https://www.hudsonlab.se>

# Mohammad Ghorbani

Community: WASP  
Senior Lecturer, Umeå University  
mohammad.ghorbani@umu.se



**Looking for open problems/questions to create new collaborations,  
Have an open project or idea that could benefit from a collaboration,  
Have a research tool or approach and want to find a collaborator with an idea or  
data to apply it to**

**Research areas: epidemiology and infection biology, AI/Math**

**Key words: Covid 19, Epidemiology, Infection disease, Markov process, Spatially  
dependent functional data, spatio-temporal statistics**

## Large scale analysis of spatio-temporal dynamics of Covid 19 epidemic data

In order to maintain human and animal health, it is extremely important to understand how pathogens such as viruses are transmitted and evolve to a higher degree of harmfulness. In other words, it is essential to understand the spatio-temporal point pattern structure of Covid 19 to effectively prevent and control this and future pandemics. The current studies have attempted to estimate the basic reproductive number of confirmed cases without considering the impact of the spatial pattern and mobility of infected cases, population density, climatic condition, socio-economy, and humidity on virus transmission.

As far as we know, so far no serious research has been conducted worldwide, and especially in Sweden, to understand the spatio-temporal point pattern of confirmed cases and deaths and to discover the role of effective factors on virus transmission.

For Covid 19, it is of interest to understand how the locations of infected individuals, population density, individual heterogeneity, travel patterns, climate condition, and various preventive measures collectively influence the course of the epidemic over time.

So, from a statistical and epidemiological perspectives, the scientific questions of interest are as follows:

- i) Are the spatial and temporal components of the Coronavirus epidemic dataset independent? in other words, is the spread of the Covid 19 related to the geographical locations of the infected regions or hot spots (at the county level) in Sweden?
- ii) Where and when unusual spacetime patterns occur in the coronavirus epidemic dataset?
- iii) Can the inhomogeneity pattern in the infected cases and deaths be explained by covariates, such as population density, socio-economy, humidity, underlying diseases (diabetes, lung disease, heart

disease), and smoking habits?

iv) In which counties have intervention methods helped mitigate the spread of Covid 19?

Note that questions i), ii) and iii) that complement each other are in the area of spatio-temporal statistics while question iv) is in the area of functional data analysis (FDA). Consequently, the main objectives of this study are:

a) To develop and apply exploratory statistical methods together with more advanced methods based on hierarchical spatio-temporal random field models defined either at the individual level or at the small scale level to analyse the data.

b) To modify a spatiotemporal risk function estimator for rapid detection of anomalies in the spatiotemporal point pattern of Covid 19 using Stockholm as a test case.

c) Analysis of Covid 19 in the framework of spatially dependent FDA.

d) To create an infectious disease website for the whole of Sweden based on daily information from the Swedish Public Health Agency, analyze the data and publish the results on the web with tabular, graphical, and map-based summaries, and update the results based on new information for online surveillance. This website could be very fruitful regarding future pandemics.

Briefly, the aim of this research is to construct statistically realistic models and develop inferential methods to describe the spread of infection in space and time; short-term prediction and long-term scenario forecasting of pandemic developments; evaluation and comparison of alternative interventions to answer some of the most pressing questions for Covid 19 and preparedness for other outbreaks.

## Short Bio

I am a senior lecturer in mathematical statistics, my main research interests are spatial and spatial-temporal statistics, analysis of spatially dependent function data, and machine learning.





## Notes