2026
PROJECT CATALOG
SciLifeLab Stockholm Summer Intern program

SciLifeLab

LIST OF PROJECTS

1. Nanoparticle-enabled drug delivery system against bacterial infections
Georgios Sotiriou, Karolinska Institutet (georgios.sotiriou@scilifelab.se)

2. Investigating blood-brain barrier dysfunction in schizophrenia-on-a-chip.
Thomas Winkler, KTH (winklert@kth.se)

3. Biophysics as a new hallmark of aging
Erdinc Sezgin, Karolinska Institutet (erdinc.sezgin@ki.se)

4. Spatial host-microbiome interactions through Spatial metaTranscriptomics (SmT)
Stefania Giacomello, KTH (stefania.giacomello@scilifelab.se)

5. Multi-modal spatial omics of neurodegeneration: From single cells to 1000-plex proteomics
Burcu Ayoglu, KTH (burcu.ayoglu@scilifelab.se)

6. Exploring Hormone receptors as biomarkers to predict tumor immune microenvironment and survival outcomes in lung cancer
Cecilia Williams, KTH (cecilia.williams@scilifelab.se)

7. MicroRNA discovery in fungi
Marc Friedländer, Stockholm University (marc.friedlander@scilifelab.se)

8. Amplified tRNA gene networks as hidden drivers of cancer
Claudia Kutter, Karolinska Institutet (claudia.kutter@scilifelab.se)

9. Determining codon-anticodon relationships at the single-cell level using tTEscanR
Claudia Kutter, Karolinska Institutet (claudia.kutter@ki.se)

10. Accelerating reproducible analysis of high-throughput sequencing data with osync
Claudia Kutter, Karolinska Institutet (claudia.kutter@ki.se)

11. Plasma proteomics and patient-reported symptoms for precision diagnostics in lung cancer
Janne Lehtiö, Karolinska Institutet (janne.lehtio@ki.se)

12. Modelling multi-modal responses to gene expression dosage in functional genomics data
Tuuli Lappalainen, KTH (tuuli.lappalainen@scilifelab.se)

13. Subcellular proteomics of acute myeloid leukemia (AML)
Janne Lehtiö, Karolinska Institutet (janne.lehtio@ki.se)

1.

PI: Georgios Sotiriou, georgios.sotiriou@scilifelab.se

Supervisor: Shengtao Yu, shengtao.yu@scilifelab.com

Nanoparticle-enabled drug delivery system against bacterial infections

The discovery of the first antibiotic, penicillin, saved millions of lives after nearly a century, which makes many people mistakenly believe that antibiotics are universal cure for everything. Decades of misuse and overuse have resulted in the emergence of antibiotic-resistant "superbugs", posing a serious global health challenge. Therefore, we urgently need to develop new strategies for bacterial infection treatments.

One promising approach is the rational use of antibiotics through precision drug delivery and dose reduction without compromising therapeutic effectiveness. In this context, multi-component drug delivery systems that exploit synergistic effects between different antibacterial agents are of growing interest. Our lab has previously demonstrated a new type of nanoparticles that have intrinsic antibacterial activity, which also showed the potential as drug carriers for delivering a wide range of therapeutic molecules, such as antimicrobial peptides, anti-tumor molecules, and mRNA.

The aim of this project is to investigate the use of antibacterial nanoparticles as carriers for antibacterial molecules and to evaluate whether nanoparticle-mediated delivery enhances antibacterial activity. The intern will prepare drug-nanoparticle conjugates and perform comparative antibacterial studies. The results will contribute to ongoing efforts to develop effective nano-enabled antibacterial therapies and provide preliminary data for future studies.

The list of instruments the student will learn. Transmission electron microscopy (TEM) to observe the morphology of nanoparticles and conjugates. Fourier transform infrared spectroscopy (FTIR) to confirm the molecule is loaded. Dynamic lighting scattering (DLS) will be used to measure the hydrodynamic diameter of the conjugates in the liquid media. Bioscreen will be used to monitor the bacterial growth, to evaluate the antibacterial activity of the conjugates under different concentrations. Fluorescent microscopy will also be used to evaluate antibacterial activity using Live/Dead viability kit. Overall, through this summer internship, the student will gain hands-on experience in nanomaterial characterization and antibacterial testing within an interdisciplinary research environment combining nanotechnology and microbiology.

The first week the student will read some research papers to build up basic knowledge of this project and attend lab safety training to be able to work in the lab. The second and third week, the student will work in the lab to perform drug loading experiments and optimize the loading efficiency, as well as characterize the conjugates to confirm the molecule is loaded through the instruments that are listed above. The fourth and fifth week, the student will conduct antibacterial activity evaluations, through Bioscreen and the fluorescent microscopy. The last week, the student will wrap up the project, including data analysis, preparation a short presentation, and writing a project report.

2.

Name PI: Thomas Winkler
Email PI: winklert@kth.se

Name supervisor: Mina Kazemzadeh Dastjerd
Email supervisor: minakaz@kth.se

Project title: Investigating blood-brain barrier dysfunction in schizophrenia-on-a-chip.

Project description:

Schizophrenia (SZ) is a severe psychiatric disorder with a poorly understood aetiology and no curative therapies. Having a better understanding of the cellular and molecular mechanisms driving SZ development is thus essential. To date, most research has focused on neuronal dysfunction, neglecting (interactions with) other brain cell types, particularly the blood-brain barrier (BBB)'s brain microvascular endothelial cells (BMECs) and pericytes.

The aim of this project is to investigate cell-cell interactions and cell type-specific contributions to BBB dysfunction in SZ. For this, our lab is currently establishing human iPSC-derived transwell cultures of the SZ-specific BBB, including BMECs and pericytes. To better mimic the dynamic physiological environment of the BBB, we now aim to advance this work using a microfluidic BBB-on-chip system, which enables co-culture under flow and functional assessment in a more physiologically relevant setting. The intern will work closely alongside a postdoctoral researcher to evaluate and select the most suitable of two commercially available microfluidic chip systems, and to optimize robust co-culture conditions for iPSC-derived BBB cells. The initial focus will be on establishing stable cultures using healthy control-derived cells and implementing readouts for the evaluation of the barrier integrity.

This work will provide the foundation for the development of a SZ-relevant BBB-on-chip platform for future mechanistic studies of BBB dysfunction, cell-cell interactions, and dysregulated signalling pathways in SZ.

List of used techniques:

Basic cell culture techniques; iPSC maintenance and differentiation; microfluidic chip handling; RT-qPCR; immunofluorescent cell staining; permeability assay.

Supervision plan:

The intern will be supervised by a postdoctoral researcher in the lab, receiving day-to-day guidance and feedback. Progress will be discussed in regular meetings with the postdoc and in scheduled meetings with the PI.

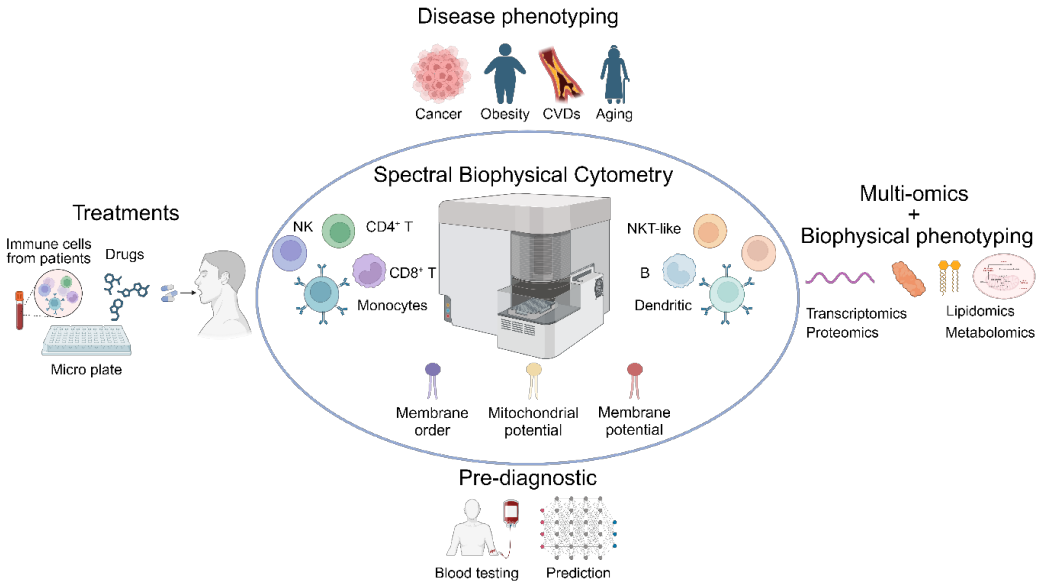
3.

Biophysics as a new hallmark of aging

PI: Erdinc Sezgin (KI/SciLifeLab) – erdinc.sezgin@ki.se
Supervisors: Cenk Gurdap (PhD student) – cenk.gurdap@ki.se;
Maria Tsiokou (PhD student) – maria.tsiokou@ki.se

Description of the project: Cells undergo biophysical remodeling as a response to environmental changes, for survival and to execute their function. Changes in the biophysical properties of cells, such as membrane fluidity, tension, and membrane potential, are crucial to various cell states or diseases and have considerable potential as a diagnostic or prognostic marker. Over the years, numerous instruments, such as atomic force microscopy, micropipette aspiration, and optical traps have been developed to investigate the biophysical characteristics of cells. However, these technologies suffer from low throughput, which is a major obstacle to applying them to medical problems that require measuring thousands of cells. Therefore, there is still a major gap in our understanding of how collective biophysical properties of the cells alter during both physiological and pathological processes. We recently described a high-throughput and multi-parametric platform with single-cell resolution, which combines spectral flow cytometry and environment-sensitive dyes. Our platform enables fast and correlative measurement of the biophysical properties of model membranes, cell lines, and human blood samples.

The student will apply the method to 1) biophysically phenotype aged cells 2) integrate biophysical phenotyping with multi-omics technologies 3) predicting biological age of cells using machine learning based on biophysical remodeling 4) explore drug treatments on aged cells using the biophysical measurements as output.



List of main techniques: Spectral flow cytometry, environmental-sensitive dyes, cell culture, molecular biology techniques, drug treatments, machine learning.

Supervision: The student will be mainly supervised by Cenk and Maria. Different parts of the project and techniques could be tailored to the student's interests. We provide training and daily supervision as needed. The student is more than welcome to engage in other lab activities, such as biophysics seminars and weekly lab meetings.

SciLifeLab

4.

PI: Stefania Giacomello, stefania.giacomello@scilifelab.se, KTH
Supervisor: Sofia Rouot (PhD student, sofia.rouot@scilifelab.se), Maria Iovinelli (PhD student, maria.iovinelli@scilifelab.se), Hailey Sounart (postdoc, hailey.sounart@scilifelab.se)

Project: Spatial host-microbiome interactions through Spatial metaTranscriptomics (SmT)

Project description

Alterations in microbiome composition and function have been associated with a broad range of disorders, including cancers such as colorectal cancer (CRC) (1), as well as inflammatory bowel disease (IBD) (2), metabolic (e.g. type 2 diabetes) (3), reproductive (e.g. endometriosis) (4) and infectious (e.g. tonsillitis) (5) diseases. These extensive associations highlight the critical need to investigate host-microbiome interactions at the cellular and molecular levels to better understand the pathophysiology of these conditions.

To this end, our research group developed Spatial metaTranscriptomics (SmT) (6), an innovative technology that simultaneously captures host and microbial 16S profiles in tissue sections, combined with matched microscopy to link spatial sequencing barcodes to the histological context. We are specifically investigating the colon tissue, to understand how factors such as disease state, dietary intake, and biological sex modulate host-microbiome niches in the colon at the spatial level.

We are looking for a motivated student with an interest in host-microbiome interactions and Spatial Transcriptomics (ST), who will learn to apply SmT on tissue samples. In parallel, the student will have the opportunity to establish a bulk metagenomics sequencing workflow and to run RNAScope from the same samples for validation purposes. Overall, the student will gain extensive knowledge and skills in SmT.

List of techniques

- SmT and ST: tissue sectioning, histological staining, imaging, qPCR, next-generation sequencing (Illumina)
- Establishing a bulk metagenomics sequencing approach
- Knowledge about the RNAScope assay

Supervision

The student will be joining the Spatial Biology group led by Assoc. Prof. Stefania Giacomello and be supervised by one of the three supervisors (Sofia, Maria, and Hailey) at all times. The student will also interact with the other members of the research group and will be able to work independently for literature review.

References

1. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012 Feb 1;22(2):299–306.

2. Joossens M, Huys G, Cnockaert M, Preter VD, Verbeke K, Rutgeerts P, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut. 2011 May 1;60(5):631–7.

3. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012 Oct;490(7418):55–60.

4. Leonardi M, Hicks C, El-Assaad F, El-Omar E, Condous G. Endometriosis and the microbiome: a systematic review. BJOG Int J Obstet Gynaecol. 2020;127(2):239–49.

5. Jensen A, Fagö-Olsen H, Sørensen CH, Kilian M. Molecular Mapping to Species Level of the Tonsillar Crypt Microbiota Associated with Health and Recurrent Tonsillitis. PLOS ONE. 2013 Feb 21;8(2):e56418.

6. Saarenpää S, Shalev O, Ashkenazy H, Carlos V, Lundberg DS, Weigel D, et al. Spatial metatranscriptomics resolves host–bacteria–fungi interactomes. Nat Biotechnol. 2024 Sep;42(9):1384–93.

5.

Multi-modal spatial omics of neurodegeneration: From single cells to 1000-plex proteomics

PI: Burcu Ayoglu, burcu.ayoglu@scilifelab.se

On-site co-supervisor: Iva Sutevski, iva.sutevski@scilifelab.se

Project background and aim: The human brain contains billions of neuronal and glial cells across hundreds of unique cell subtypes and ~200 brain regions. Neurodegenerative diseases (NDDs) disrupt this highly organised system at multiple levels - from protein misfolding and neuronal loss, to synaptic dysfunction, impaired energy metabolism, cytoskeletal disorganisation, and neuroinflammation. Despite decades of research, the protein-level relationships between these pathological processes remain poorly understood, and as a result, no disease-modifying treatments are currently available.

Our project focuses on major NDDs - Alzheimer's disease (AD), Lewy body diseases (LBDs), and Frontotemporal dementia (FTD). Although these disorders start in distinct brain regions and present with distinct early symptoms, they spread to neighbouring brain regions and increasingly overlap in clinical representation. We aim to identify **disease-specific pathological signatures in human brain tissue** using **spatial omics**.

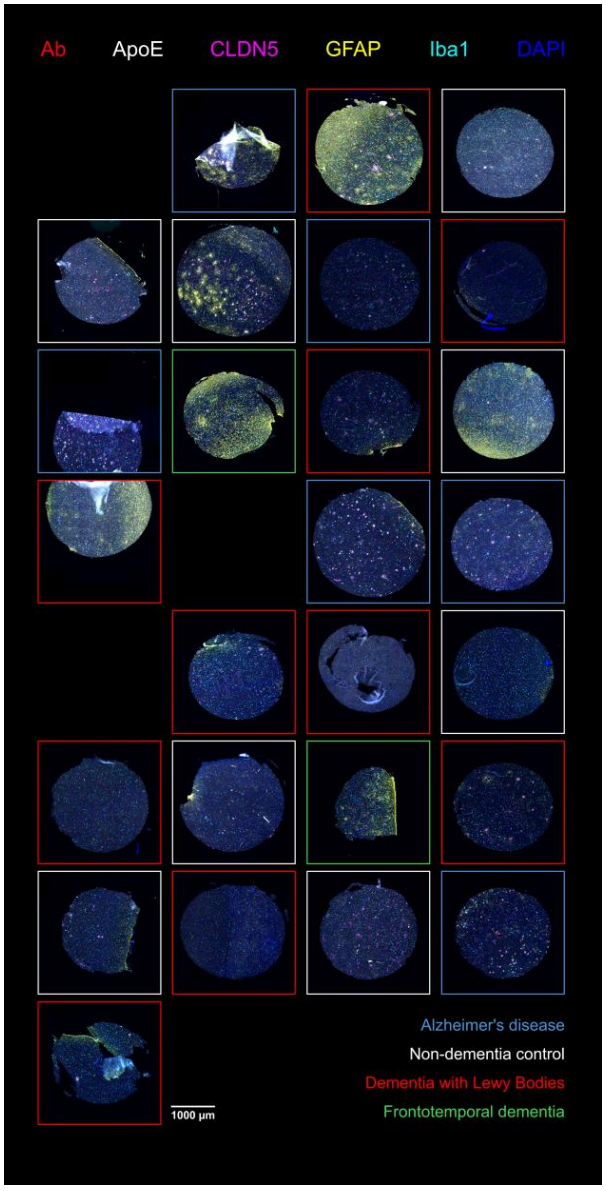
We integrate three cutting-edge spatial omics approaches applied on consecutive brain tissue sections:

1) PhenoCycler-Fusion for automated cyclic immunofluorescence imaging up to ~100 proteins at subcellular resolution (~0.5 µm), enabling precise cell segmentation as well as neighbourhood and morphology analyses,

2) GeoMx Discovery Proteome Atlas (DPA) platform, enabling us to profile deeper up to ~1200 proteins within selected "regions of illumination", and finally

3) StereoSeq, a spatial transcriptomics approach.

By combining single-cell resolution imaging with ultra-high-plex protein profiling and transcriptome-level profiling, we generate a multi-layered, spatially resolved molecular atlas of neurodegeneration, linking cellular architecture to large-scale protein expression patterns.



Representative staining patterns from a 42-plex multiplex immunofluorescence analysis of a custom human brain tissue microarray we constructed, comprising 27 grey matter cores from three NDD groups and controls.

SciLifeLab

5.

What you will learn:

During this summer internship with us, you will gain hands-on experience with:

- **Generating single-cell resolved multiplex immunofluorescent microscopy data with the PhenoCycler Fusion** platform
- **Generating ultra high-plex immunofluorescent microscopy data using the GeoMx DPA** platform
- **Computational image processing and spatial omics data analysis** using in-house developed tools and custom Python scripts

By the end of this internship, you will have acquired a comprehensive understanding of the full workflow-from sample preparation to data acquisition and analysis-involved in two different, cutting-edge spatial proteomics approaches. Furthermore, you will be involved in optimising data integration generated from three different spatial omics modalities. Finally, you will also apply these state-of-the-art techniques to advance our understanding of tissue biology in different neurodegenerative diseases.

6.

Exploring Hormone receptors as biomarkers to predict tumor immune microenvironment and survival outcomes in lung cancer

PI: Cecilia Williams, cecilia.williams@scilifelab.se
Project co-supervisor: Baizhen Chen, baizhen.chen@scilifelab.se

Project plan

Cancers are heterogeneous and influenced by several factors such as age, heredity, lifestyle, metabolic disorders, and inflammation. Growing evidence highlights sex-specificity in cancer development with hormones such as estrogen acting either as protective or pro-tumorigenic factors.
Colorectal cancer (CRC) shows specific sex differences with men exhibiting an earlier onset, a higher incidence, and a lower survival compared to age-matched pre-menopausal women. We have demonstrated that estrogen signaling via the estrogen receptor β (ERβ/ESR2) protects against neoplastic lesions formation and modulate the colonic immune microenvironment in a sex-dependent manner. Its expression is lost in cancer cells, however,an analysis of TCGA CRC data reveals improved survival in patients expressing higher ERβ trace level with notable sex-specificity. Single-cell transcriptomic analysis indicates that ERβ expression in tumor is mainly driven by defined immune cell populations. We have also observed that the expression of other hormone receptors (HR) such as ERα and the androgen receptor (AR) are correlated with differential survival in CRC.

We hypothesize that hormone receptor expression in certain cancers could be linked to specific immune signatures in the tumor microenvironment and may be used as predictive biomarker. Our preliminary data show that in kidney cancer high ERβ expression correlates with poor survival prognostic for both men and women. In contrast, in lung cancer, high ERβ expression together with low ERα expression correlates with improved survival in men specifically.

The proposed bioinformatics project aims to investigate the link between sex, hormone receptors expression, tumor immune-microenvironment and survival outcomes in lung cancer patients. This project is part of a broader PAN-cancer analysis framework to explore hormone receptors as potential biomarkers for cancer prognosis.

Supervision plan

The intern will be co-supervised by PhD student Baizhen Chen under a 6-week period. The intern will develop their own bioinformatics pipeline around a pre-established structure as a reference. The project will be divided into 5 tasks.
Task 1 (week1): Introduction to the project. Literature research and research of publicly available datasets for lung cancer (we have already acquired some). Discussion about the analysis strategy.
Task 2 (week 2): Correlation and survival analysis. High and low expression groups (overall patients) for selected hormone receptors will be determined. Data from KM curves and from Cox-regression model will be compared. The groups will be also separated to investigate sex-difference.
Task 3 (weeks 3-4): Differential gene and gene ontology analysis. Transcriptomic analysis using public bulk RNA-seq data for high vs low expression groups for selected hormone receptors.
Task 4 (weeks 5-6): Deconvolution of bulk RNA-seq data will be performed to explore changes in the immune cell composition between High and low expression groups.

List of techniques

R-program
RNA-seq analysis (depending on the experience of the intern scRNA-seq analysis can be considered)
Differential gene analysis (DEseq2)
Gene ontology analysis (ClusterProfiler)
Survival analysis (KM curve and Cox-regression model)
Bulk-RNA-seq deconvolution using CIBERSORTx

SUMMER STUDENT PROJECT: microRNA DISCOVERY IN FUNGI

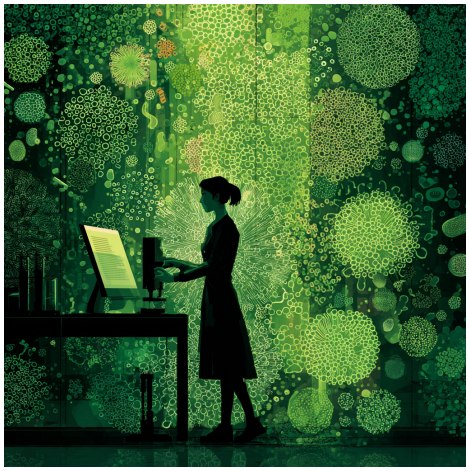
PI: Marc Friedländer (marc.friedlander@scilifelab.se)

Project supervisors:
Marc Friedländer (marc.friedlander@scilifelab.se)
Valentin Poltorachenko (valentin.poltorachenko@scilifelab.se)

Project description:
MicroRNAs are molecules around 22 nucleotides long that regulate the expression of protein coding genes and play important roles in numerous biological processes, including development, stress responses and organismal health. MicroRNAs are found in all organisms with complex multicellularity, including plants and animals. Surprisingly, although fungi have all the proteins necessary to produce these molecules, no microRNAs have so far been discovered in these microbes. To our knowledge, the search for microRNAs have only been undertaken in a few fungal model organisms, such as baker's yeast, and no systematic studies have been conducted.

As part of an independent ERC consortium effort, our collaborator Vicent Pelechano has obtained 100 distinct fungal species in his lab. We plan to use a split-and-pool approach to sequence small RNAs from these 100 fungal species using next-generation sequencing. The summer student will then map the sequenced RNAs to the individual fungal genomes and apply our in-house miRDeep2 software (Friedländer et al., Nucleic Acids Research 2012) to discover microRNAs in these genomes. This is a very high-risk high-gain project that - if successful - will report the first microRNAs ever found in fungi, and will rewrite how we think of gene regulation in microbes.

Techniques:
The summer student will learn dry-lab methods related to small RNAseq analysis. This will include quality control with miRTrace (Kang et al., Genome Biology 2018), mapping using bowtie1 and various genomes and annotations, microRNA discovery using miRDeep2 and curation of a shortlist of the discovered candidates.



AI graphics by Midjourney

Supervision:
The project will be supervised by Marc Friedländer (myself) who is a world-leading expert in microRNA discovery (in all humility). The project will be co-supervised by PhD student Valentin Poltorachenko. Vicent Pelechano will advise on the fungal genomics aspects of the study. The student will be supervised by weekly meetings in person and when not possible in select weeks via zoom.

8.

1. Project subject: Amplified tRNA gene networks as hidden drivers of cancer

2. Responsible PI and supervisor:

Claudia Kutter (group leader, claudia.kutter@ki.se),
Keyi Geng (day-to-day supervisor, Postdoctoral researcher, keyi.geng@ki.se)

3. Project description:

Background: In cancer genetic alterations disrupt normal regulatory programs and drive uncontrolled proliferation and therapy resistance. Most studies have focused on protein-coding genes and mRNAs, but transfer RNAs (tRNAs) remain understudied. Traditionally viewed as adaptor molecules in translation, tRNAs are now recognized as active regulators of cancer biology. Cancer cells may increase the dosage of specific tRNA genes, thereby expanding the cellular tRNA pool and directly reshaping the translational machinery. Such a mechanism could provide a selective advantage by boosting translation of codon-biased oncogenic mRNAs and enhancing cellular fitness under stress or drug treatment.

Objectives: This student project will examine how amplified tRNA genes in cancer cells influence transcription and cell behaviour. We will (**i**) map their genomic organization and co-occurring gene modules using sequencing and computational tools, (**ii**) assess transcriptional activity and localization of tRNA genes, and (**iii**) test their functional impact on cell proliferation, stress response, and drug sensitivity through cell-based assays.

4. Techniques:

Cell-based assays:

- Cell culturing of various human cancer cell lines
- Proliferation and viability assays
- Drug sensitivity and stress response assays
- Confocal microscopy for DNA/RNA localization
- Measurement of translational efficiency

Molecular techniques:

- CRISPR/Cas-mediated perturbations (overexpression, knock-down)
- Pol III ChIP-seq and/or ATAC-seq
- RNA-seq after cell perturbation

Computational techniques:

- Computational codon-anticodon usage integration analysis (tTEscanR or alike)
- Sequencing data analysis (ChIP-seq, ATAC-seq, RNA-seq)
- Visualization of genomic and transcriptomic data

Project management skills:

- Reproducible research practices (ELN)
- Weekly group meetings, journal club participation, and final report preparation

5. Central References:

1. Kutter et al. Nature Genetics, 2011 https://www.nature.com/articles/ng.906
2. Schmitt et al. Genome Research, 2014 https://genome.cshlp.org/content/24/11/1797.full
3. Rudolph et al. PLoS Genetics, 2016 https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006024
4. Gao et al. Genome Research, 2022 https://genome.cshlp.org/content/32/1/97.long

9.

1. Project subject: Determining codon-anticodon relationships at the single-cell level using **tTEscanR** (**t**heoretical **T**ranslation **E**fficiency **scan**ner implemented in **R**)

2. Responsible PI and supervisor:

Claudia Kutter (group leader, claudia.kutter@ki.se),
Ana Varas Sanchez (day-to-day supervisor, PhD student, ana.varas.sanchez@ki.se)

3. Project description:

Background: Translation efficiency depends on the balance between codon demand (mRNAs) and anticodon supply (tRNAs). While previous studies showed codon-anticodon balance across species, organ development, and cell states, bulk analyses masked cell-to-cell variability. Single-cell RNA-seq and ATAC-seq now allow investigation of translational efficiency in individual cells, revealing subsets with unbalanced pools that create translational bottlenecks or selectively enhance protein synthesis.

Objectives: This project will use the tTEscanR pipeline, developed in the Kutter group, to quantify codon and anticodon usage at single-cell resolution in disease contexts, including cancer, neurological disorders, and viral infections using publicly available data. The student will map codon demand and tRNA supply, calculate cell-type-specific translation efficiency, and explore how translational programs differ across cell types or disease states.

4. Techniques:

Computational:

- Single-cell RNA-seq and ATAC-seq data processing
- Codon and anticodon quantification with tTEscanR
- Integration of codon-anticodon data to estimate translational efficiency
- Data visualization and interpretation
- Optional: pipeline improvement and software development

Data management skills:

- Reproducible coding practices and version control (GitHub)
- Quality control and preprocessing of single-cell datasets
- Weekly group meetings, journal club participation, and final report preparation

Experimental (optional, if time permits):

- Ribo-seq in mouse and human samples

5. Central References:

1. Kutter et al. Nature Genetics, 2011 https://www.nature.com/articles/ng.906
2. Schmitt et al. Genome Research, 2014 https://genome.cshlp.org/content/24/11/1797.full
3. Rudolph et al. PLoS Genetics, 2016 https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006024
4. Gao et al. Genome Research, 2022 https://genome.cshlp.org/content/32/1/97.long
5. https://github.com/wgao688/sc_tRNA_mRNA

10.

1. Project subject: Accelerating reproducible analysis of high-throughput sequencing data with *osync* (*o*mics *sync*hronization)

2. Responsible PI and supervisor:

Claudia Kutter (group leader, claudia.kutter@ki.se),
Carlos Gallado Dodd (day-to-day supervisor, PhD student, carlos.gallardo.dodd@ki.se)

3. Project description:

Background: Modern biology is driven by massive sequencing datasets, from single-cell RNA-seq to ATAC-seq and beyond. These high-throughput datasets are key to understanding development, disease, and evolution, but analyzing them reliably is challenging. Standard workflows are often complex, error-prone, and difficult to reproduce. *osync* is a cutting-edge computational framework, developed in the group, that automates sequencing pipelines, ensures reproducibility, and scales efficiently across datasets of any size. By streamlining multi-step analyses, *osync* allows researchers to focus on the biology rather than troubleshooting data pipelines.

Objectives: This MSc project will enable the student to use *osync* for reproducible and high-performance processing of sequencing data analysis. Key goals include:

- Mastering *osync* workflows and understanding pipeline automation for diverse sequencing datasets.
- Processing large-scale RNA-seq, ATAC-seq, or ChIP-seq datasets efficiently and reproducibly (linked to an ongoing research project in the group)
- Benchmarking *osync* performance against traditional tools for speed, reliability, and reproducibility.
- Optionally, extending the framework with new modules or features to enhance usability, visualization, or automation.

The student will gain hands-on experience with state-of-the-art computational pipelines, work with large-scale genomic datasets, and contribute to tools that accelerate biomedical research.

4. Techniques:

Computational and analytical skills:

- Learn and master *osync* and Snakemake workflows for sequencing pipelines
- Preprocessing and quality control of sequencing data
- Alignment and quantification (STAR, Bowtie2, featureCounts)
- Integrative analysis and visualization in R or Python
- Reproducible research practices (Git/GitHub, pipeline documentation)
- Optional: pipeline extension or module development

Project Practices:

- Handling and managing large-scale sequencing datasets
- Weekly progress meetings, literature review, and data presentation
- Final report and reproducible pipeline demonstration

5. Central References:

1. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012;28:2520–2522.
2. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017;35:316–319.
3. Grüning B, Chilton J, Köster J, et al. Practical computational reproducibility in the life sciences. Cell Syst. 2018;6:631–635.

11.

Plasma proteomics and patient-reported symptoms for precision diagnostics in lung cancer

PI: Janne Lehtiö, janne.lehtio@ki.se
Supervisor: Noora Sissala, noora.sissala@ki.se
Affiliation: Department of Oncology-Pathology, Karolinska Institutet

Project description

Early detection of lung cancer remains a major clinical challenge, and there is growing interest in minimally invasive tools that could support diagnosis at an earlier stage. Modern proteomics technologies now make it possible to measure thousands of proteins from a small blood sample, offering new opportunities to identify molecular markers of disease. Meanwhile, patients report a wide range of symptoms, sometimes long before diagnosis, but these data are rarely collected systematically or linked to molecular data. In this project, we aim to combine plasma proteomic measurements with detailed symptom information to both investigate the biological processes underlying symptom development and to develop diagnostic models for early detection of lung cancer.

The dataset includes plasma proteomics and detailed symptoms data from 114 patients referred to the Karolinska University Hospital with suspected lung cancer. Patients' plasma samples have been analyzed using global mass spectrometry-based proteomics and Olink® Explore, providing quantification of >4,000 proteins. The symptoms data includes >180 distinct symptom descriptors divided into 11 symptom modules (e.g. cough, pain, fatigue).

Student role & possible objectives

This project offers a valuable opportunity for a master student to gain hands-on experience in analyzing high-dimensional proteomics and symptoms data from a real clinical cohort, while contributing to interdisciplinary research at the intersection of molecular biology and patient care. The student will apply computational methods (preferably in R) to analyze, integrate, and interpret the data. Depending on the student's interest, the project may focus on:

- Protein-symptom associations
    - Explore correlations between proteins and specific symptoms or symptom severity
    - Identify biological pathways associated with key symptoms (e.g., fatigue, cough, weight loss)
- Symptom clustering and molecular signatures
    - Identify clusters of co-occurring symptoms
    - Identify proteomic profiles associated with specific symptom patterns
    - Compare protein-symptom associations between cancer and benign disease

11.

- Integrative diagnostic modeling
    - Train and evaluate machine learning models to distinguish lung cancer from benign disease based on proteomics, symptoms, and clinical data
    - Explore different models and data integration strategies

The student will be jointly supervised by a PhD student (main supervisor) and PostDoc (co-supervisor) at Lehtiö lab (https://lehtiolab.github.io/).

Learning outcomes

- Understanding of plasma proteomics technologies
- Experience in preprocessing and quality control of high-dimensional omics data
- Skills in statistical and computational analysis
    - Correlation, regression, and clustering analyses
    - Machine learning modeling for classification problems
    - Multimodal data integration
- Coding in R (and/or possibly Python)
- Data visualization and scientific communication
- Insight into translational cancer research

12.

Modelling multi-modal responses to gene expression dosage in functional genomics data

Project PI: Tuuli Lappalainen, tuuli.lappalainen@scilifelab.se
Project Supervisor: Leah Rosen, leah.rosen@scilifelab.se

Project Description:

How cells respond to quantitative changes in gene expression is a vital question across fields, and in recent years we finally have the tools to study this, using both functional genomics (e.g. CRISPRi/a and exogenous overexpression) and population genetic approaches. In the Lappalainen lab we have generated single cell CRISPR screen data to understand transcriptomic responses to quantitative changes in gene expression, while also developing a bayesian model, called bayesDREAM (Bayesian modelling of Dosage Response Effects Across Modalities) to model dosage response curves from these data. BayesDREAM works on data with diverse modalities and shows excellent power and sensitivity. Having meaningful dose-response curves from bayesDREAM allows us to tackle a diverse set of problems.

- Problem 1: While it is known that transcription factor dosage determines cellular outcomes during development, differentiation, reprogramming, and disease progression, there is much that we still don't understand about these dosage dynamics. Applying bayesDREAM to differentiation data will help us understand how changes in dosage both cause differentiation, and how differentiation changes dosage response curves.
- Problem 2: Current gene regulatory network models rely on CRISPR screen data, but make the simplifying assumption that dosage response curves are linear. We can apply bayesDREAM-fitted dosage response curves to enhance existing gene regulatory network models.
- Problem 3: Given that over 90% of human trait and disease-linked genetic variants have been found in the non-coding genome with many of these variants quantitatively affecting gene expression, understanding gene dosage response is particularly relevant in human genetics. We're hoping to apply bayesDREAM to human population data.

Which of the three problems the student chooses to tackle will depend on their interests and background.

Techniques used:

The student should be proficient in either python or R. In particular, they will be handling large datasets, visualising results, and working with existing packages. The student will perform Bayesian modelling in python (using pyro), and modify existing package code, appropriate to their interests and experience. The student will work on an HPC cluster environment, involving handling package installs and managing environments.

Supervision plan:

Leah Rosen, a postdoc in Tuuli Lappalainen's lab will provide day-to-day supervision, with regular meetings with Professor Tuuli Lappalainen, who will guide the project directions and help with biological interpretation.

13.

Project title: Subcellular proteomics of acute myeloid leukemia (AML)

Principal Investigator (PI): Prof. Janne Lehtiö, janne.lehtio@scilifelab.se, KI
Supervisor: Yanbo Pan (PhD, Senior Research Specialist), yanbo.pan@scilifelab.se, KI

Objectives

- Generate subcellular proteomic profiles of AML cell lines.
- Identify proteins with significant re-localization linked to drug response.
- Provide hands-on experience in mass spectrometry-based proteomics and bioinformatics.

Project description

Protein localization plays a vital role in cellular function, and its misregulation can contribute to disease progression and drug resistance in AML. **Subcellular proteomics** has proven to be a powerful tool in identifying disease mechanisms by tracking protein distribution changes[1].

We have developed **SubCellBarCode** [2-4], which offers a straightforward method for robust protein localization and re-localization analysis for more than 15000 proteins. This technique has been successfully applied to study protein re-localization in response to various cellular perturbations, such as EGFR inhibition and protein farnesylation inhibition. By expanding its application to AML cells treated with anti-leukemia drugs, we aim to gain insights into drug mechanisms of action, resistance, and off-target toxicity.

Here, we are looking for a motivated student with an interest in subcellular proteomics. The student will learn to apply the SubCellBarCode method on AML cells with/without drug treatment, to explore the protein locations and translocations at a proteome-wide level. The SubCellBarCode method is a combination of laboratory work and bioinformatic analysis, we expect the student to work in a well-documented and reproducible way and the focus of the student will be the adaption of well-developed SubCellBarCode method on AML cells to generate the new knowledge to understanding leukemia. The project will provide the student with many new insights into MS-based proteomics, spatial proteomics, bioinformatic pipelines, as well as the involved leukemia biology.

Key techniques the student will use and learn

- Protein subcellular fractionation strategies for subcellular proteomics.
- Sample preparation for MS analysis, such as cell culturing, protein extraction/digestion.
- MS-based quantitative proteomics including both Data Independent Acquisition (DIA) and Tandem Mass Tagging (TMT) labeling methods.
- Reproducible computational subcellular proteomics analysis by **SubCellBarCode** R package.

Supervision

- Daily support from Yanbo Pan and senior lab member.
- Weekly meetings with Yanbo Pan to track progress, troubleshoot challenges, and refine methods.
- Mid-internship evaluation to assess progress and adjust goals if needed.
- Final review session with Yanbo Pan and Prof. Janne Lehtiö to discuss findings.

Deliverables

- A processed dataset with annotated subcellular localization and re-localization.
- A short scientific report/presentation summarizing methods, findings, and interpretations.

Reference

1. Josie A Christopher et al. ***Nat Rev Methods Primers***. 2021, 1:32.
2. Lukas Orre[#], Mattias Vesterlund[#], Yanbo Pan[#], et al. ***Mol Cell***. 2019, 3;73(1):166-182.e7.
3. Taner Arslan[#], Yanbo Pan[#], et al. ***Nat Protoc***. 2022, 17(8):1832-1867.
4. Yanbo Pan[#], et al. ***iScience*** 2025, doi: https:// doi.org/10.1016/j.isci.2025.111864.

SciLifeLab