2023
PROJECT CATALOG
SciLifeLab Stockholm Summer Intern program

SciLifeLab

LIST OF PROJECTS

1. **RNA Folding by Deep Learning**
   Samuel Flores (samuel.flores@scilifelab.se), Stockholm University

2. **Codon-driven translational efficiencies across diseases at single cell level,**
   Claudia Kutter (claudia.kutter@ki.se), SciLifeLab/KI-MTC

3. **Modelling actin mesh reorganisation at the T-cell immunological synapse,**
   Juliette Griffie (juliette.griffie@dbb.su.se), SU

4. **Does T-cell molecular organisation encode for biological sex?**
   Juliette Griffie (juliette.griffie@dbb.su.se), SU

5. **Spatial proteomic profiling of childhood leukemia,**
   Janne Lehtiö (janne.lehtio@scilifelab.se), Karolinska Institutet,
   Supervisor: Yanbo Pan, yanbo.pan@scilifelab.se

6. **PCR-based neural networks for composition-sensitive diagnostics,**
   Ian Hoffecker (ian.hoffecker@scilifelab.se), KTH,
   Supervisor: Erik Benson, erik.benson@scilifelab.se

7. **Transcriptional effects of acute and chronic SSRI administration on brain networks,**
   Iskra Pollak Dorocic (iskra.pollak@scilifelab.se), Stockholm University,
   Supervisor: Charlotta Henningson, charlotta.hennings@scilifelab.se

8. **Expanding KEGG and Reactome pathways with a network-based approach,**
   Erik Sonnhammer (erik.sonnhammer@scilifelab.se), Stockholm University (DBB),
   Supervisor: Davide Buzzao, davide.buzzao@scilifelab.se

9. **Evaluation of alternative data formats for genomic sequencing data,**
   Anja Mezger (anja.mezger@scilifelab.se), National Genomics Infrastructure (NGI); KTH Royal Institute of Technology,
   Supervisor: Matthias Zepper, matthias.zepper@scilifelab.se

10. **Measuring content and biophysical properties of extracellular vesicles,**
    Erdinc Sezgin (erdinc.sezgin@scilifelab.se), Karolinska Institutet

11. **Bringing together what belongs together - Imputing spatial information from single-cell transcriptome data,**
    Vicent Pelechano (vicent.pelechano@scilifelab.se), Karolinska Institutet (MTC) and SciLifeLab,
    Supervisor: Marcel Tarbier, marcel.tarbier@scilifelab.se

1.

RNA Folding by Deep Learning

PI: Samuel Coulbourn Flores, samuel.flores@scilifelab.se

AlphaFold2 has revolutionized structural biology, solving the structure of most proteins based on sequence. However RNA structure prediction is much more difficult, due to higher charge, size, and flexibility, less training data, and a smaller alphabet. Past Summer Fellow Chih-Fan Chang made a significant breakthrough, helping to classify tetraloops with much more accuracy than expected. However the accuracy should be boosted further by inclusion of longer sequences and multiple alignments. We welcome an additional MS student to contribute to this exciting project.

Techniques: information theory, RNA folding by internal coordinates, deep learning

Supervision plan: individual or small-group meetings at least weekly, plus group meetings.

2.

1. Project subject: Codon-driven translational efficiencies across diseases at single cell level

2. Cutting-edge methods in life sciences: single cell RNA-seq, single cell ATAC-seq, Ribo-seq in mouse and human

3. Background: The Central Dogma connects the three molecules found in all life: DNA, RNA, and protein. Transcription directs information from DNA to RNA and translation from RNA to protein. Translation involves communication between messenger RNAs (mRNAs) and transfer RNAs (tRNAs). Thus, translational efficiency is affected by both the codon demand of mRNAs and the anticodon supply of matching tRNAs. We previously showed that the codon-anticodon pools are balanced during species evolution[1], organ development[2] and in differentiating vs. quiescent cells[3]. These studies were limited because we compared a bulk of cells (e.g., whole liver) rather than individual cells. Technological advances and the availability of single cell atlases allowed us to assess whether a subset of cells within a large population use unbalanced pools of codons to control translational efficiencies[4]. By using these atlases, we confirmed that the codon-anticodon pools are highly stable across mammalian cell types, but certain cell types have unbalanced codon usage that creates a translational 'choke' point, while others exhibit increased supply of certain anticodons to enhance translation efficiency[4].

4. Project tasks: This *computational* project will leverage on the expertise, methods, and tools (GitHub scripts) developed and available in the Kutter group to investigate the codon-anticodon relationship in diseases (e.g., cancer, neurological disorders, viral infections etc). The specific tasks for the student to be complemented within 8 weeks will be:

Tasks	**Weeks**							
(research and learning activities)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Familiarize with research topic (literature)								
Familiarize with in house methodology and GitHub scrips[5]								
Familiarize with scATAC-seq perturbation data								
Learning code control for reproducible research								
Identifying complementing scRNA-seq and scATAC-seq data								
Data quality control								
Quantification of codon usage and amino acid demand								
Quantification of anticodon usage and amino acid supply								
Measure of translation efficiency at cell type level								
Analysis of complementing data based on prior findings								
Visualization and data interpretation								
Optional: software development								
Keeping meticulous records								
Self-education (literature, coding)								
Weekly one-to-one meetings								
Participation at group meeting and journal club								
Handover of data and final report								

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

6. Supervisors: Claudia Kutter (PI), Carlos Gallardo Dodd (PhD student)

3.

Supervisor: Juliette Griffie – juliette.griffie@dbb.su.se

Computational microscopy for cell biology laboratory @SciLifeLab

Affiliated: Stockholm University

Modelling actin mesh reorganisation at the T-cell immunological synapse.

Project description:
Actin is part of the cell cytoskeleton. It has been shown to play a fundamental role in T-cell signalling: from activation to migration (https://www.nature.com/articles/ni0700_23). Relying on super resolution microscopy, it is now possible to visualise its nanoscale organisation and remodelling over key cellular processes. Strikingly, there is no available simulation platform dedicated to generating realistic synthetic data sets of the actin mesh (and its remodelling) at the T-cell immunological synapse. It is a considerable drawback for the development of new analytical tools. This research project will tackle this challenge. It focuses on developing a novel simulation platform for actin mesh synthetic data sets during T cell activation, uniquely enabling for live cell features such as flow or depolymerisation to be incorporated. It is a computational project with deeply rooted biological applications.

List of techniques: Computational project relying on building a biophysical model for actin filaments, as well as coding (in python ideally), spatial point pattern strategies and tracking.
Supervision: The student will be joining the Computational Microscopy for Cell Biology (CMCB) lab, led by Assitant Prof. J. Griffié (https://www.scilifelab.se/researchers/juliette-griffie/). He/She will be meeting weekly with the supervisor, as well as joining CMCB group meeting and socials. She/He will also benefit from training and technical support (both on the computation side and the biological context) throughout the project.

4.

Supervisor: Juliette Griffie – juliette.griffie@dbb.su.se

Computational microscopy for cell biology laboratory @SciLifeLab

Affiliated: Stockholm University

Project: Does T-cell molecular organisation encode for biological sex?

Project description:
Super resolution (SR) microscopy encompasses advanced light microscopy techniques which have in common to circumvent the diffraction limit (Nobel Prize in Chemistry 2014). More concretely, it uniquely enables to image fluorophores in cells with a resolution ranging from a few nm to 200nm. It provides unprecedented insight into cells' molecular life and has highlighted the key role played by signalling proteins spatial (re)organization and interactions at the nanoscale. As such SR has opened novel promising avenues of research when it comes to investigating complex cellular processes.
T-cells play an essential role in the build-up of immune responses, and SR provides an outstanding platform to improve our understanding of their activation at the molecular level. Although there are known clinical differences between biological sexes when it comes to auto immunity, very little is known on how it translates in terms of the spatio-temporal distribution of key signalling proteins during T-cell activation. This research project will focus on imaging, with SR technologies, T cells before and during activation to investigate if biological sex information is encoded at the nanoscale using cutting edge analytical tools developed by the host lab. Visualising, quantifying and in fine understanding these subtle differences, brings us a step closer to personalised medicine. It is a highly multidisciplinary research project, relying on cutting edge imaging and analysis strategies to answer a fundamental biological question.
List of techniques: Super resolution microscopy, Single molecule localisation microscopy (https://www.nature.com/articles/s43586-021-00038-x), T-cell artificial synapses, immuno-staining, cluster analysis tools and co-clustering (https://www.sciencedirect.com/science/article/pii/S2211124720315126).

Supervision: The student will be joining the Computational Microscopy for Cell Biology (CMCB) lab, led by Assitant Prof. J. Griffié (https://www.scilifelab.se/researchers/juliette-griffie/). He/She will be meeting weekly with the supervisor, as well as joining CMCB group meeting and socials. She/He will also benefit from experimental and computational training throughout the project.

5.

PI: Prof. Janne Lehtiö, janne.lehtio@scilifelab.se, KI
Supervisor: Yanbo Pan (Ph.D., Researcher), yanbo.pan@scilifelab.se, KI

Project: Spatial proteomic profiling of childhood leukemia

Project description

Protein localization is important in many aspects of cellular function and can also play a role in the development and progression of leukemia. In childhood leukemia, changes in protein localization can be associated with drug resistance and disease progression. Mass spectrometry (MS)-based spatial proteomics has provided important insights into the relationship between protein function and subcellular location[1]. By mapping protein subcellular location and re-localization events caused by cellular perturbations, spatial proteomics will be able to uncover important information on leukemia pathophysiology, drug mechanism of action, and off-target toxicity in leukemia cells, and eventually overcome drug resistance and improve treatment outcomes for children with leukemia.

We have developed SubCellBarCode [2,3], which offers a straightforward method for robust protein localization and re-localization analysis. We have demonstrated the power of the SubCellBarCode in a proteome-wide analysis of protein re-localization in response to EGFR inhibition. The possibility of investigating condition-dependent localization will enable uncovering the molecular response to a wide range of perturbations and add important new knowledge in the disease research field.

Here, we are looking for a motivated student with an interest in subcellular proteomics. The student will learn to apply the SubCellBarCode method on leukemia 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 main focus of the student will be the adaption of well-developed SubCellBarCode method on leukemia cells to generate the new knowledge to understanding childhood 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. Maybe it is out of the scope of a summer intern project, but we hope we can use the generated data here to expand the existing project “SubCellBarCode 2.0” in our group.

List of techniques

- Cell line culturing, drug treatment, and protein subcellular fractionation;
- Sample preparation for MS-based quantitative proteomics, such as protein extraction/ digestion, tandem mass tag (TMT) labeling;
- Knowledge about liquid chromatography (LC) and mass-spectrometry: mechanism and applications;
- Reproducible computational analysis and basic bioinformatic analysis by R/Python.

Supervision

The project will be supervised by Researcher Yanbo Pan, who is responsible for the SubCellBarCode project in Prof. Janne Lehtiö group. The student will be continuously supervised by Yanbo Pan throughout the entire internship. The generated data would be immensely helpful to the group and should integrate with the existing project of SubCellBarCode.

Bibliography

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.

6.

PCR-based neural networks for composition-sensitive diagnostics

PI: Ian Hoffecker, ian.hoffecker@scilifelab.se
Deputy supervisor: Erik Benson, erik.benson@scilifelab.se

Project description
PCR (polymerase chain reaction) has revolutionized biology and recently played an important role in the form of high sensitivity detection for diagnostics amidst the ongoing Covid19 pandemic. While a clear binary (yes/no) detection of the presence or absence of a molecule is a useful basic design for a diagnostic, often a system's complex composition and the relative proportions of a mixture of different molecules is the only available insight into the pathogenicity of a sample. For example the relative expression levels of certain genes may determine if a tissue is healthy or cancerous, and in both cases, the same set of genes might be expressed with only their relative proportions differing, rendering a binary detection diagnostic ineffective. The goal of this project is to test the feasibility, via a working PCR simulation framework that we have developed previously, of a composition-sensitive PCR diagnostic based on the idea of a DNA-based neural network. That is, rather than implementing a neural network in classical computing hardware, we would be simulating the possibility of representing nodes and edges in the neural network with entirely DNA sequence-based material. The simulation framework will make use of sequence-specific DNA hybridization physics to simulate the differential interaction between strands according to their sequences that would enable the assignment of weights, and we would program the relative concentrations and reaction conditions to control bias. The student would get to experience the development of an exciting and completely novel technology, gain broadly applicable exposure to the biophysics of DNA, and get hands on experience in bioinformatics coding in the powerful and rising Julia programming language.
Techniques:

- Programming and bioinformatics in the Julia/python programming languages
- Data analysis and visualization
- Neural networks
- DNA hybridization physics and polymerase chain reaction

Supervision
The student will receive a focused introduction to the topic in the first week and thereafter work semi-independently with weekly meetings with the group leader. The student will also be encouraged to work with more senior lab members for additional guidance.

7.

Transcriptional effects of acute and chronic SSRI administration on brain networks

Research group/PI: Iskra Pollak Dorocic, Stockholm University, iskra.pollak@scilifelab.se

Supervisor: Charlotta Henningson, charlotta.hennings@scilifelab.se

Project description:

Neuropsychiatric disorders are among the most common but also least understood health problems. Mood disorders, including major depression, have high lifetime prevalence, are associated with substantial morbidity, and cause significant burden both to the individual and to society. Current treatments have significant limitations in terms of efficacy and side effects, and progress in finding new therapies has been slow.

Selective serotonin reuptake inhibitors (SSRIs) are the first-line treatment for major depression disorder, and among the most prescribed drugs in the world. However, the precise molecular and transcriptional effects of SSRI administration on their primary target serotonin neurons are not known. This project studies the effect of acute and chronic administration of the SSRI fluoxetine to determine the transcriptional changes in the brain following treatment. We hypothesize that SSRI administration will cause specific transcriptional changes in different neural populations, including serotonin and likely non-serotonin neurons.

The student will work with a spatial transcriptomic data set and use a variety of analysis pipelines and statistical methods to perform bioinformatics. The student will work independently on a defined research question, but will be supervised by a PhD student who is chiefly responsible for this project and will be able to guide the progress.

Our lab's overall aim is to identify novel serotonin neuron subtypes based on molecular profile and eventually circuit anatomy, and link them to a functional role which may be impaired in various psychiatric disorders. The student will also have the opportunity to learn about in-vivo approaches used in the lab to study brain connectivity, neural activity and behavior.

Methods used:

- Analysis of spatial transcriptomics sequencing data
- Multiplex data integrations with RNAscope in-situ results
- Additional integration of public single cell RNA seq data sets
- Advanced statistics
- Adaptation of analysis pipelines and custom scripting in R

8.

Research plan - Summer interns 2023

Expanding KEGG and Reactome pathways with a network-based approach

Stockholm University (DBB)
Supervisor: Prof. Erik Sonnhammer
Co-supervisor: Ph.D. student Davide Buzzao

Project description

FunCoup is a web-based network biology resource (Alexeyenko and Sonnhammer 2009; Persson et al. 2021) that is intended to help researchers in identifying and examining functionally coupled genes/proteins. The platform is a useful tool for biologists doing systems biology since it offers a user-friendly interface for interpreting and displaying the results.

A genome-scale biological network like FunCoup could be used to fill the gap of high incompleteness that popular pathway databases such as KEGG (Kanehisa et al. 2014) and Reactome (Croft et al. 2011) suffer from (Gable et al. 2022). As shown in Gable et al., both KEGG and Reactome are characterized by a low genome coverage and gene annotation biases. With a supervised clustering approach, KEGG and Reactome pathways may be enhanced and these expanded pathways can then be used for improved pathway analysis. In this project, the FunCoup subnetwork for each KEGG pathway will be extracted and then extended with TOPAS (Buzzao et al. 2022) and MaxLink (Guala, Sjölund, and Sonnhammer 2014). The extent to which new pathway members were added will then be assessed, as well as the quality of the extended pathways by comparing to the Gene Ontology. Writing scripts (preferably in R or Python) and results analysis will be required for this project. The student will present to the group and prepare a final report.

Bioinformatics databases and tools to be used

- Online data sources:
  - Networks from FunCoup https://funcoup.sbc.su.se/;
  - Pathways from KEGG https://www.kegg.jp/ , Reactome https://reactome.org/
- Offline programmatic methods:
  - Network visualization with iGraph (python, R), NetworkX (python);
  - Network module analysis with TOPAS https://bitbucket.org/sonnhammergroup/topas/ and MaxLink https://maxlink.sbc.su.se/download/

9.

Evaluation of alternative data formats for genomic sequencing data

Matthias Zepper: Project-supervising bioinformatician, NGI Genomic Applications: matthias.zepper@scilifelab.se

Johannes Alneberg: Bioinformatics Lead, NGI Applications Development: johannes.alneberg@scilifelab.se

Anja Mezger: Head of Facility, NGI Genomic Applications: anja.mezger@scilifelab.se

Introduction

Inside a cell, genetic information is usually well stored: Tightly wrapped around histones, arranged on chromosomes, possibly packed up in a nucleus. Unfortunately, once we at the National Genomics Infrastructure sequenced it, this no longer applies unconditionally.

However, cells also had a head start of several hundred million years to figure out how to properly handle genetic information and we can improve, too.

Project description

Essentially, commonly used data formats for genomic sequencing data are compressed text files,[1;2] which can be used outright in many applications. However, they are rather bulky, don't allow for straightforward access to subsets of the data and are not searchable.

Over the past years, several alternative data formats, dedicated compression algorithms and indexes have been devised,[3;4;5;6;7] but haven't found widespread adoption.

The ever increasing amounts of sequencing data, however, suggest to rethink the status quo, in particular for archival of unprocessed data. With analysis workloads moving to public clouds, new requirements emerge. When object storage rather than file systems is used, access patterns will change and new stream-based, selective access methods may become beneficial.

As technology provider, the NGI strives to provide users with meaningful, innovative, yet also mature solutions. Therefore, we think it is time to subject the current developments to a thorough test.

Project goals and milestones

- Gather and select tools to be tested by literature search.
- Compile a project plan and define suitable test criteria, such as compression rate and computational effort.
- Test the various tools, data formats and methods on high-performance compute infrastructure.
- Summarize results and key findings in a presentation at the SciLifeLab. Optionally, design a scientific poster.

Guidance and supervision will be provided as needed and preferred.

Various stretch goals are possible, such as writing a conversion pipeline in Nextflow or fuzzing the published conversion softwares.

Applicant profile

This is a purely bioinformatic project with a strong focus on data engineering rather than biological applications. We welcome a quick learner, who is interested to acquaint themselves with some details of data formats and compression algorithms. Familiarity with the command line interface is required. Prior coding experience, particularly in compiled languages such as C++ or Rust, is favorable.

Acquirable skills and experiences

Data formats and structure ◇ compression algorithms ◇ job execution on high-performance computing systems ◇ data visualization and presentation ◇ data analysis pipelines ◇ CLI tools ◇ version control

References

[1] Peter J. A. Cock, Christopher J. Fields, Naohisa Goto, Michael L. Heuer, and Peter M. Rice. The sanger fastq file format for sequences with quality scores, and the solexa/illumina fastq variants. *Nucleic acids research*, 38:1767–1771, April 2010. ISSN 1362-4962. doi: 10.1093/nar/gkp1137.

[2] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and samtools. *Bioinformatics (Oxford, England)*, 25:2078–2079, August 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp352.

[3] Faraz Hach, Ibrahim Numanagić, and S. Cenk Sahinalp. Deez: reference-based compression by local assembly. *Nature methods*, 11:1082–1084, November 2014. ISSN 1548-7105. doi: 10.1038/nmeth.3133.

[4] Prashant Pandey, Fatemeh Almodaresi, Michael A. Bender, Michael Ferdman, Rob Johnson, and Rob Patro. Mantis: A fast, small, and exact large-scale sequence-search index. *Cell systems*, 7:201–207.e4, August 2018. ISSN 2405-4712. doi: 10.1016/j.cels.2018.05.021.

[5] Sultan Al Yami and Chun-Hsi Huang. Lfastqc: A lossless non-reference-based fastq compressor. *PloS one*, 14:e0224806, 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0224806.

[6] Divon Lan, Ray Tobler, Yassine Souilmi, and Bastien Llamas. Genozip - a universal extensible genomic data compressor. *Bioinformatics (Oxford, England)*, 37: 2225–2230, February 2021. ISSN 1367-4811. doi: 10.1093/bioinformatics/btab102.

[7] Robert Bakarić, Damir Korenčić, Dalibor Hršak, and Strahil Ristov. Sfq: Constructing and querying a succinct representation of fastq files. *Electronics*, 11(11), 2022. ISSN 2079-9292. doi: 10.3390/electronics11111783.

10.

Measuring content and biophysical properties of extracellular vesicles

PI: Erdinc Sezgin
Description of the project
Extracellular vesicles (EVs) are key components in cell-cell communication. Recently, their key roles in health and disease were revealed. However, it is still challenging to study them due to their small size (<100nm). We recently developed a single particle profiling technology that enables us to study nano-sized bioparticles such as EVs.
In the proposed project, we aim to use our new technology to study EV content and biophysical properties. We will study whether EVs from different sources (cancer cells, blood etc), and different preparation methods give rise to different EVs. Moreover, the properties we will measure will shed light on the source of EVs in the body and give us insight on how to use them as efficient delivery agents.

Key papers and techniques:
Single particle profiler:
Single particle profiler for measuring properties of nano-sized bioparticles
Taras Sych, Florian Weber, Jan Schlegel, R. Beklem Bostancioglu, Hanna M.G. Barriga, Kariem Ezzat, Herbert Stangl, Birgit Plochberger, Samir El Andaloussi, Molly M. Stevens, Jurga Laurencikiene, Erdinc Sezgin
bioRxiv 2022.07.08.499323; doi: https://doi.org/10.1101/2022.07.08.499323
Biophysical tools:
Dissecting the mechanisms of environment sensitivity of smart probes for quantitative assessment of membrane properties.
Ragaller F, Andronico L, Sykora J, Kulig W, Rog T, Urem YB, Abhinav, Danylchuk DI, Hof M, Klymchenko A, Amaro M, Vattulainen I, Sezgin E.
Open Biology, 2022 Sep;12(9):220175. doi: https://doi.org/10.1098/rsob.220175
Synthetic biology:
Influence of the extracellular domain size on the dynamic behavior of membrane proteins.
Gurdap CO, Wedemann L, Sych T, Sezgin E.
Biophys J. 2022 Oct 18;121(20):3826-3836. doi: https://doi.org/10.1016/j.bpj.2022.09.010

Supervision:
We have open door policy in the lab. The student will be supervised by Erdinc Sezgin as well as postdocs and PhD students in the lab. We will also have collaborating labs at KI.

11.

Bringing together what belongs together

Imputing spatial information from single-cell transcriptome data

PI: Vicent Pelechano (KI, MTC)
Supervisor: Marcel Tarbier (postdoc)

Project description:

Single-cell sequencing has revolutionized our understanding of cellular heterogeneity, but since it relies on tissue dissociation it erases all spatial context. It is, however, well understood that gene expression is not just a function of a cell's state but also integrates signals from its spatial context, the so-called micro-environment. So without spatial information we cannot fully understand molecular mechanisms in complex tissue environments, which in turn limits our understanding of, e.g., cancer recurrence or tissue regeneration.

High-throughput approaches to study spatial patterns of the entire transcriptome do not reach single-cell resolution, or are limited to nuclear RNA and lack sensitivity. Existing approaches to integrate of 'spatial transcriptomics' data with matched single-cell sequencing data are promising, but fail to assign a single-cell to an exact spatial context. Therefore, to study the impact of the micro-environment on the single-cell level, one is currently limited to low-throughput imaging approaches, e.g., in-situ sequencing or sequential FISH.

We are therefore developing computational approaches to impute spatial context purely based on single-cell RNA sequencing data. For this we integrate spatial data with single-cell transcriptomics data to learn subtle patterns that are indicative for different cellular neighborhoods. Initial analyses show great promise, and this summer will therefore be the perfect time for a student to join our efforts to disentangle the relationship between gene expression and spatial context.

Our lab offers extensive experience in quantitative single-cell analysis as well as data integration. We are looking for a student who has a good understanding of statistics and is experienced in coding with R (experience with omics-technologies comes handy as well). The student will have the opportunity to work independently, receive close supervision as desired or needed, to develop and test their own ideas, and to develop skills in spatial biology, single-cell data analysis and data integration.

- Analysis of single-cell RNA sequencing data
- Analysis of spatial transcriptomics / in-situ sequencing data
- Multi-omics data integration
- Advanced statistics
- Custom scripting in R