2022
PROJECT CATALOG
SciLifeLab Stockholm Summer Intern program

SciLifeLab

Maria Pernemalm, maria.pernemalm@scilifelab.se
KI

1.

Supervisor: Haris Babačić, haris.babacic@scilifelab.se

Project: ***Plasma proteome alterations in glioblastoma***

Project Description:

Glioblastoma (GBM) is the most frequent and most malignant primary brain tumour, characterised by a very short survival after diagnosis, usually not longer than 15 months due to lack of therapeutic options. Currently, there are few biomarkers that are clinically useful in predicting response to treatment and survival in GBM and all of them require tissue sampling for analysis. Due to the location of the tumour, it is difficult to access tumour tissue for biopsy. This frequently requires opening the patient's skull even in situations when the tumour is inoperable. To bypass this need, we aim to identify circulating protein biomarkers in the liquid part of the blood (i.e., plasma) that can predict survival and response to treatment in patients with GBM. Furthermore, we aim to demonstrate that the blood-brain barrier is impaired in GBM and proteins can leak into the blood.

For this purpose, we have performed the most in-depth plasma proteome profiling by mass-spectrometry proteomics to date, using a method developed in our lab. We have already identified over 3,000 plasma proteins in GBM patients, and several proteins that may serve as biomarkers in GBM. During the summer project, the student will be involved in running another cohort of GBM patients, to validate previously discovered biomarkers.

Techniques taught:

The student will learn how to prepare plasma samples for mass-spectrometry based proteomics analysis with high-resolution isoelectric focusing (HiRIEF) coupled to liquid chromatography (LC) and mass-spectrometry (MS-MS). The student will be introduced into the field of plasma proteomics and taught how to perform plasma depletion of highly-abundant proteins, measure protein concentration, run polyacrylamide gel electrophoresis (PAGE), and digest proteins into peptides. Should the student be successful, the time might allow for learning how to perform labelling with tandem mass tags (TMT) and HiRIEF. Learning high-dimensional statistics and analysis of proteomics data in R is optional, depending on the student's interests.

Project Plan:

Week(s)	Tasks
1	Introduction to mass-spectrometry proteomics and plasma proteomics Introduction to the lab and lab safety.
2-4	Lab work: Plasma depletion and buffer exchange
4-6	Lab work: Protein concentration measurements and PAGE
6-8	Lab work: Protein digestion. Optional: TMT-labelling and HiRIEF.

Samuel Flores, samuel.flores@scilifelab.se
Stockholm University

2.

Project: ***RNA 3D Structure Prediction by Deep Learning***

AlphaFold2 has revolutionized structural biology, solving the structure of most proteins based on sequence. However RNA structure prediction is well behind this, as its unique physical characteristics, reduced alphabet, and much smaller amount of available training data, mean it cannot be solved in precisely the same way. We have devised a novel approach to this problem which we are now busy applying. 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.

SciLifeLab

Anders Andersson (anders.andersson@scilifelab.se)
Department of Gene Technology, KTH / SciLifeLab

3.

Project: ***Reconstructing fungal genomes from soil metagenomes***

Project description:

Soil-associated fungi play key roles in environmental cycles and represent both threats and opportunities in face of climate change, as they can both contribute to release of climate gasses under warming conditions, and contribute to carbon sequestration in symbiosis with plants. Fungi are an integral component of natural soil microbial communities. However, excluding pathogens with major medical or economic relevance, their lifestyle and genetic potential are not well described, and few fungal genomes are sequenced. Consequently, the genetic capacity of fungi in increasing soil resilience and reducing the impact of climate change remains largely unexplored. Therefore, the overall goal of the project is to assemble genomes of uncultivated fungi from the publicly available shotgun metagenomics dataset of global top soil samples (Bahram *et al., Nature* (560) 2018), with an adapted version of the computational method CONCOCT (Alneberg *et al, Nat Methods* (11) 2014).

Shotgun metagenomics generates sequences of all DNA present in a sample. The method is widely used for investigating the genetic potential of microbial communities. In addition, it allows reconstruction of genomes from complex communities, without the need for isolation and cultivation of the organisms. For genome reconstruction, overlapping sequencing reads will be assembled into contigs, and the contigs derived from the same organism binned into genomes (Metagenome-Assembled Genomes; MAGs). The A.Andersson group successfully developed the now widely used algorithm CONCOCT for binning contigs into MAGs based on the contigs' sequence composition and abundance distribution across multiple samples (Alneberg *et al, Nat Methods* (11) 2014). The method was established for small prokaryotic genomes, and has not yet been extensively evaluated for larger eukaryotic genomes. In this project CONCOCT will be applied to bin fungal contigs into genomes using soil metagenome data. We will compare running assembly and binning directly, with an approach where eukaryotic contigs are preselected based on their k-mer profile before the binning step (West *et al, Genome Res* (28) 2018). The approaches will be evaluated on a subset of 40 samples, and the best approach used on the extensive dataset of global topsoil samples (Bahram *et al., Nature* (560) 2018), which will allow us to assemble genomes of uncultivated fungi and investigate their genetic potential in processes of carbon cycling and sequestration.

List of techniques:

- Bioinformatics (e.g UNIX, python, R)
- Data analysis & visualisation (e.g. R, Python)
- Literature search
- Work on computing cluster (UNIX)
- Create analysis workflow (Snakemake)

Supervision:

The student's project is linked to a project of postdoc Meike Latz and planned to result or be included in a publication. The student will be supervised by Meike Latz and Anders Andersson.

Tuuli Lappalainen, tuuli.lappalainen@scilifelab.se
KTH

4.

Supervisor: Philipp Rentzsch, philipp.rentzsch@scilifelab.se
Project: ***Generation of genome-wide conservation scores based on Zoonomia alignments***

Project description

One of the foundations for understanding molecular biology is that functionally important genomic regions are evolutionarily conserved between related species. Based on this assumption, genomic conservation scores like phastCons (Siepel et al. 2005), phyloP (Pollard et al. 2010), and GERP (Davydov et al. 2010) have long been used to study the human genome. These scores are generated by analyzing every single position in a multiple species genome alignment with a statistical test. The derived metrics have been found to be highly correlated with functional constraint and can be used to predict individual variant effects or rank and prioritize causal positions in a quantitative trait loci (QTL).
Recently, the Zoonomia Project has released unprecedented multiple species alignments of 240 different mammalian species. In the associated paper (Zoonomia Consortium 2020), the authors have exemplary generated the conservation score phyloP based on these alignments and show that they can explain many previously uncharacterized genomic elements. We are very much interested in these findings. Here, we are looking for a student that is doing a similar kind of analysis based on the conservation scores GERP and phastCons.
For this project, we are seeking a motivated student with an interest in computational data analysis. The project does not require a strong background in software development or statistics (though that obviously cannot hurt). Your main focus will be the adaption of an already existing data analysis to the new, bigger data set. We expect you to work in a well-documented and reproducible way. Depending on your own interest and prior knowledge, the project will provide you with many new insights in HPC computing, software pipelines, statistics, as well as the involved biological disciplines.
While likely out of the scope of an eight-week project, we ultimately want to use the generated conservation scores to support existing projects in our group in characterizing functional genomic regions (regulatory elements, enhancer & splicing QTLs, GTEx).

List of techniques student will use/learn

- reproducible computational analysis and documentation
- parallel data processing in an HPC environment (Uppmax)
- handling of large (TB+) data files
- development/integration of domain knowledge of Multiple-Sequence-Alignments, Genome Reference Sequences and Genome Conservation

What type of supervision will the intern have during the 8 weeks?

The project will be supervised by postdoc Philipp Rentzsch, who has done similar analyses in the past. As is, the project is is "stand alone" and does not depend on other projects. However, generated results would be immensely helpful to the group and should integrate with (and support) many already existing analyzes.

5.

Validation and characterisation of molecules binding ATXN2, a candidate target in drug discovery against Amyotrophic Lateral Sclerosis

Principal Investigator: Prof. Oscar Fernandez-Capetillo
Affiliation: Karolinska Institutet, Department of Medical Biochemistry and Biophysics, Division of Genome Biology

The lab

Our group focuses on the use of cell-based phenotypic screens to identify new molecules of medical interest. In a bottom-up strategy, we combine high-throughput genetic and chemical screening methods with classical cell- and molecular-biology techniques. This comprehensive approach starts from molecule discovery followed by the mechanistic elucidation of its properties. Our scientific interest is on age-related diseases lacking efficient cures such as cancer or neurodegeneration.

The project

One of the interests of our group is Amyotrophic Lateral Sclerosis (ALS), a fatal neurodegenerative disease characterized by the loss of motor neurons. Given the low incidence of ALS, there is little incentive in the pharmaceutic industry to prioritize drug development in this area, making academic efforts even more important. In one of ALS-related projects, we are focused on finding molecules that can perturb the function of ATXN2, which was shown to alleviate ALS pathology using genetic models. Unfortunately, ATXN2 is not an easy drug target due to its lack of enzymatic activity. We have therefore taken advantage of a novel imaging-based method developed at SciLifeLab (*in situ* CETSA) to look for molecules that bind to ATXN2 and could be further modified to degrade the protein specifically. Having found several candidate molecules, the following steps are to validate and characterize them. The student's project will be focused on the characterization of these new drugs, their properties, and their capability to impair ATXN2 function.

Techniques

The student will use a wide range of techniques including cell culture, Western blotting, thermal shift assay, immuno-precipitation, immunocytochemistry, high-throughput high-content microscopy and automated live-cell imaging. Data analysis will encompass automated image analysis, plotting and statistical analyses of big data.

Supervision

Besides the overall supervision of PI, the student will be continuously supervised by a Postdoctoral researcher involved in this project throughout the entire internship. After an initial hands-on introductory period, the student will be encouraged to strive for semi-independence in planning and executing the work.

6.

The chicken or the egg?

Assessing gene expression inertia through single-cell multi-omics (scRNAseq and scATACseq) and cellular barcoding over multiple generations

PI:

Vicent Pelechano (KI, MTC), vicent.pelechano@scilifelab.se

Supervisor:

Marcel Tarbier (Post-doc), marcel.tarbier@scilifelab.se

Project description:

One of the most puzzling and most discussed questions in epigenetics is whether gene expression induces chromatin accessibility or whether the causality is reverse. To answer this question we performed single-cell multiomics - single-cell RNA sequencing for gene expression combined with single-cell ATAC-sequencing for chromatin accessibility - on a homogeneous population of mouse embryonic stem cells in collaboration with the Suter lab at EPFL in Switzerland. Additionally we trace cell relationships with RNA barcodes, so we can discern which cells are closely related to one another (e.g. being sister cells or cousins).

We just are starting the analysis of this data and summer will be the perfect time for a student to join our efforts to disentangle the relationship between gene expression, chromatin accessibility and cell lineages. Our lab has experience in quantitative single-cell analysis as well as single-cell RNAseq and ATACseq analysis and multiomics. We are looking for a student who has a good understanding of epigenetics and omics-technologies as well as statistics, and is experienced in coding with R. The student will have the opportunity to work independently, receive closer supervision when needed or desired, to develop and test their own ideas, and to develop skills in single-cell data analysis and data integration.

- Analysis of single-cell RNA sequencing data
- Analysis of single-cell ATAC sequencing (chromatin accessibility) data
- Multi-omics data integration
- Additional integration of public data sets and data bases
- Advanced statistics
- Custom scripting in R

SciLifeLab

Per-Olof Syrén, per-olof.syren@scilifelab.se
KTH

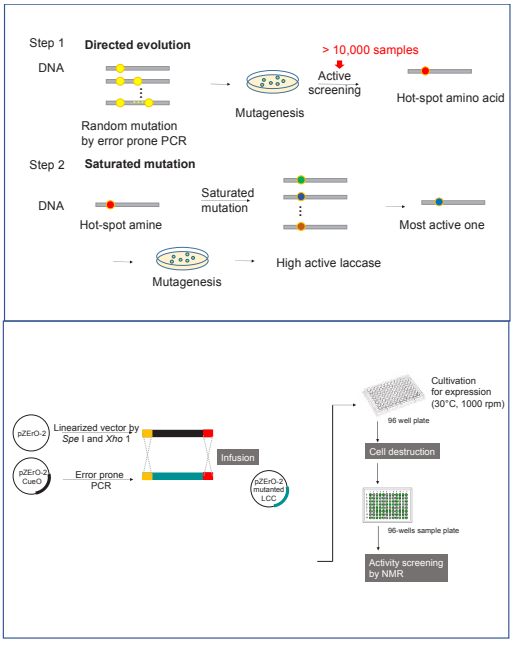
Supervisor: Boyang Guo, Boyangg@kth.se

7.

Direct evolution of plastic degrading enzyme for sustainable recycling of nylon products

Project Background and objective:
Nylon is a polyamide plastics that encompasses a broad range of market applications, including fishing nets, clothing, carpets, and packaging sheets for manufacturing. Large amounts of post-consumer nylon products are discarded every year and most of this waste currently ends up in the landfill, taking around 40 years to self-decompose, thus contributing to the global problem of plastic pollution. Nylon can be mechanically recycled; however, the steps are tedious, require significant time under harsh conditions, and cannot efficiently recycle nylon from nylon-copolymer products. Being able to recover monomers from nylon waste can reduce the need for petroleum-based resources by providing a closed-loop materials economy. **Enzymatic recycling breaks plastic polymers down to their building blocks (monomers) or a mixture of oligomers, which can then be separated from other polymer types left in the mixture. These functional substances can be processed into new products and converted to new plastics or plastic resins or other applications. However, the key difficulties in the field include the compressed structure of polymers that limit enzyme accessibility and the low polymer degrading activity of naturally existing biocatalysts.** In this work, we will take advantage of microwave technology combined with enzyme directed evolution to generate a novel variant of a plastic degrading enzyme (cutinase) that is capable of cleaving polyamides..

Project description:
To achieve the goal of enzymatically hydrolyzing nylon requires two parallel project tracks. One project is to develop nylon pre-treatment methods before the biocatalytic reaction to pre-depolymerise the bulky nylon into oligomeric chains for improved enzyme accessibility. We have successfully produced his oligomeric nylon and will use this as the substrate for positive candidate cutinase screening. Another part is focused on computational engineering of key residues in cutinase that are likelyrelevant to affect the nylon hydrolysis activity in cutinase. However, this method only targets a small area near the activity sites of the enzyme. Compared to the rational design, natural enzyme evolution randomly introduces error in the entire DNA fragment, which offers more chance to generate several mutations that significantly change the enzyme conformation and structure that could better interact with nylon. This experimental process will be your main focus and isshowed in the figure on right. We believe this project will contribute to both fundamental research and industrial applications in the bio-degradation of nylon and other plastics.



What you learn:

- In this project you will learn fundamental knowledge about biotechnological plastic degradation
- You will learn how to clone, express and proteins with *E. coli*.
- Directed evolution, how to purify protein using different methods.
- Enzyme engineering and reaction mechanisms.
- HPLC and NMR analysis.

SciLifeLab

Vicent Pelechano (vicent.pelechano@scilifelab.se)
Karolinska Institute (MTC) / SciLifeLab

8.

Supervisor: Yerma Pareja Sánchez - yerma.pareja@scilifelab.se

Project: ***Relationship between RNA turnover and growth rate in human cells***

Supervisor: Yerma Pareja Sánchez

Project description:

Changes in gene expression are generally measured based on RNA abundance. However, the total level of an RNA in the cell is the result of the balance between synthesis and degradation. Thus, the synthesis or degradation of a transcript can increase without affecting its abundance if the antagonistic process is increased in a similar extend. In that case, though, the RNA molecules will be renewed more often, the transcript will have a faster turnover. Such, faster turnover of RNA has been related with the ability of the cells to rapidly change their transcriptome and, therefore, better cope with environmental challenges. In the eukaryotic organism *Saccharomyces cerevisiae*, RNA turnover positively correlates with growth rate. This is, the faster the cell grows, the faster it renews its transcriptome. In order to investigate whether this relationship could hold for human cells, we performed metabolic labeling and parallel sequencing of total and newly synthesized RNA fractions (TT-seq) of cells growing at different rates.

The proposed project will focus on establishing the relationship between growth rate and RNA turnover in human cells using computational approaches. First, the student will analyze the RNA turnover in the conditions provided in the experimental dataset, at global and at the single gene level. The conclusions derived from the analysis will be validated by integration with data from external databases. The student will have the opportunity to work on cutting-edge approaches and to contribute with their own ideas. For this, they must have a solid understanding of statistics and modeling, and need to be confident in using R and in exploring external data sets. Likewise, the student should have general knowledge of RNA biology. The student will learn to apply advanced statistics to answer biological questions and will work in an international and interdisciplinary environment.

Sean Rudd, sean.rudd@scilifelab.se
Karolinska Institutet

Project Title – Development and characterization of small molecule inhibitors against the drug resistance factor SAMHD1

Project Site/PI – SciLifeLab Solna, group of Assistant Professor Sean Rudd (Karolinska Institutet)

Project Supervisor – Si Min Zhang, Ph.D.

Project Details

Sterile α motif histidine/aspartic acid domain containing protein 1 (SAMHD1) is a deoxynucleoside triphosphate (dNTP) triphosphohydrolase. Whilst it is normally central to cellular dNTP pool homeostasis, our previous work has demonstrated that SAMHD1 can also metabolize and deactivate antimetabolite drugs critical for the treatment of leukemias (e.g. cytarabine), leading to poor patient response and survival. More importantly, our group has shown targeting SAMHD1, through protein degradation or deactivation, drastically improves the efficacy of anti-leukemic drugs, suggesting a viable way to improve the current treatment for leukemia. However, no SAMHD1 specific, cell-active small molecule inhibitors have been reported so far.

We have conducted biochemical and cell-based high-throughput screening campaigns and thereby identified a promising lead compound TH4168 for the development of SAMHD1 inhibitors. TH4168 inhibited SAMHD1 activities in biochemical assays, bound to SAMHD1 in cells, and most excitingly, improved efficacies of anti-leukemic drugs in AML cell lines. Here in this project, with the help of a SciLifeLab Summer Intern, we aim to further our understanding of the mechanism of action of TH4168 through several biophysical techniques. Specifically,

1. Characterize interaction modality between TH4168 and SAMHD1 *in vitro* using assays such as differential scanning fluorimetry (DSF).
2. Determine and characterize interaction modalities between TH4168 and SAMHD1 in live cells through cellular thermal shift assay (CETSA).

Student in this project will be supervised by Dr. Si Min Zhang from the team of Ass. Prof. Sean Rudd, associated to the group of Professor Thomas Helleday, where a multidisciplinary group of biochemists, molecular and cell biologists, medicinal chemists, pharmacologists, and medical doctors, all work with collective focus upon improving anti-cancer therapy. In this project, the student will use an array of state-of-the-art methods, from biochemical/biophysical assays with recombinant protein, through to cell-based assays in pre-clinical cancer models, all of which are in-house reported/validated. Additionally, they will receive training in data analysis software, data presentation, and academic writing.

The project is primarily planned with the following timeline.

Week	Task	Supervision
Week 1	Lab orientation/introduction	Dr. Sean Rudd/Dr. Si Min Zhang
Week 2-7	Experiment planning, data acquisition/analysis	Dr. Si Min Zhang
Week 7-8	Final report/presentation	Dr. Si Min Zhang/Dr. Sean Rudd

For further information on our group, see https://www.scilifelab.se/researchers/sean-rudd/

SciLifeLab

Per Ljungdahl, per.ljungdahl@scilifelab.se
Stockholm University

Association of the *Candida albicans* Proline Dehydrogenase (Put1) with mitochondrial electron transporting complexes (ETC)

PI of the Study: Prof. Per O. Ljungdahl

Applicant (if different from PI): Dr. Fitz Gerald S. Silao

Background:

Candida albicans is the most important fungal pathogen of humans. Our laboratory is actively investigating how nutrients, such as the amino (imino) acid proline, modulate the virulence of this pathogen. In *C. albicans*, proline utilization occurs in the mitochondria where the enzymes proline dehydrogenase (PRODH; Put1) and 1-Pyrroline-5-carboxylate dehydrogenase (P5CDH; Put2) are localized (Fig. 1). Proline is first converted by Put1 to P5C, which is then spontaneously converted to L-glutamate γ-semialdehyde (GSA). GSA is then used by Put2 as a substrate to generate glutamate, which is then postulated to exit the mitochondria into the cytosol where the enzyme glutamate dehydrogenase (Gdh2) is localized. Strains lacking *PUT1* or *PUT2* are completely unable to utilize proline as a sole nitrogen source and have reduced virulence in several infection models such as macrophage (1), *Drosophila*, and mice (unpublished data) reinforcing the idea that proline is an important energy source for the fungal cell, especially under nutrient-limited conditions. The reduced cofactors ($FADH_2$ and NADH) generated by proline catabolism are oxidized in the mitochondria to generate ATP.

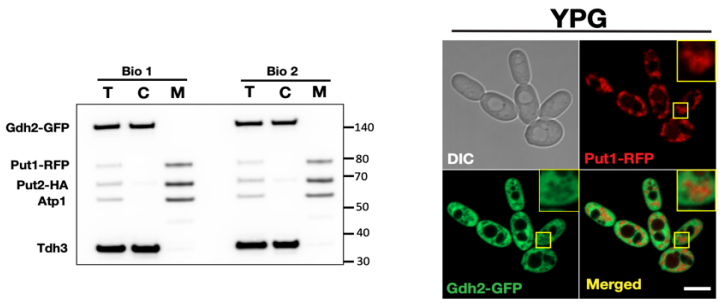


Fig. 1. Subcellular localization of proline catabolic enzymes. (Left) Strain CFG433 expressing Put1-RFP, Put2-HA, and Gdh2-GFP was pre-grown in YPDextrose overnight and then inoculated to YPGlycerol at OD_{600} = 2. Cultures were incubated with shaking at 37°C for 4 h and total (T), cytosolic (C) and mitochondrial (M) fractions were analyzed by immunoblot. Target proteins were simultaneously detected using an optimized antibody cocktail for both primary (α-GFP (1:3000), α-mcherry (RFP)(1:6000), α-Tdh3 (1:5000), α-Atp1(1:2000)) and secondary antibodies (α-mouse-HRP (1:15000), α-rabbit-HRP (1:15000), α-HA-HRP (1:15000)). Put1-RFP (~81.9 kDa) and Put2-HA (~69.8 kDa) are localized to mitochondria as they co-fractionate with the mitochondrial marker Atp1 (Alpha subunit of the F1 component; F_1F_0 ATP synthase, ~59 kDa). Gdh2-GFP (~141 kDa) is cytosolic, co-fractionating with the cytosolic marker Tdh3 (GAPDH, ~35.8 kDa). (Right) Confocal microscopy images of cells co-expressing Put1-RFP and Gdh2-GFP grown under similar condition as in left.

In eukaryotes including *C. albicans*, the catabolism of one molecule of proline yields approximately 30 ATP equivalents (2, 3), which makes it an excellent energy source. Put1 is induced in the presence of exogenous proline via the Put3-transcription factor (4). In mammalian cells, PRODH (Put1 ortholog) associates with the inner mitochondrial membrane where it interacts with the succinate dehydrogenase complex (complex II) (5). This suggests that the oxidation of proline may be directly coupled to the reduction of ubiquinone (CoQ), the known electron acceptor of the oxidation of succinate.

Aim:

This aim of this project is to determine whether the *C. albicans* Put1 protein, a proline-inducible, mitochondrial-localized enzyme, directly interacts and associates with components of the mitochondrial electron transport chain.

Supervision/Key methods:

The MS student will work together with a senior postdoctoral associate Fitz Gerald Silao. In the course of this work, the student will acquire hands-on experience with these laboratory skills:

Per Ljungdahl, per.ljungdahl@scilifelab.se
Stockholm University,

- Microbiological culture techniques
- Mitochondrial isolation and purification
- Protein Immunoprecipitation (IP/Co-IP)
- SDS-/ Blue Native (BN)-PAGE/Western blot
- Molecular cloning/CRISPR gene

Bibliography:

1. Silao, F. G. S., Ward, M., Ryman, K., Wallstrom, A., Brindefalk, B., Udekwu, K., and Ljungdahl, P. O. (2019) Mitochondrial proline catabolism activates Ras1/cAMP/PKA-induced filamentation in Candida albicans. *PLoS Genet* **15**, e1007976
2. Zhang, L., and Becker, D. F. (2015) Connecting proline metabolism and signaling pathways in plant senescence. *Front Plant Sci* **6**, 552
3. Liang, X., Zhang, L., Natarajan, S. K., and Becker, D. F. (2013) Proline mechanisms of stress survival. *Antioxid Redox Signal* **19**, 998-1011
4. Tebung, W. A., Omran, R. P., Fulton, D. L., Morschhauser, J., and Whiteway, M. (2017) Put3 Positively Regulates Proline Utilization in Candida albicans. *mSphere* **2**
5. Hancock, C. N., Liu, W., Alvord, W. G., and Phang, J. M. (2016) Co-regulation of mitochondrial respiration by proline dehydrogenase/oxidase and succinate. *Amino Acids* **48**, 859-872