**1. Functional glycan analysis in single-cell data using deep learning**

**Supervisor:** Dr Daniel Bojar, Department of Chemistry and Molecular Biology, University of Gothenburg, [daniel.bojar@gu.se](mailto:daniel.bojar@gu.se)

Glycans or complex carbohydrates are a widespread and diverse biological sequence, playing key roles in biological processes as wide-ranging as regulating the immune system or facilitating infection by pathogens. Glycomics, the study of all glycans in a sample, currently exists only as a bulk measurement. Newly developed methods use DNA-tagged lectins, proteins specifically binding to carbohydrate substructures, to measure glycan features of single cells in conjunction with single-cell transcriptomics data. In prior work (https://pubmed.ncbi.nlm.nih.gov/36217547/), we have shown that one of these glycan features can be predicted, via deep learning, from the single-cell transcriptome. Model interpretation analyses have then provided us with insights into functional roles of this glycan feature. In collaboration with the group of Dr. Hiroaki Tateno at the National Institute of Advanced Industrial Science and Technology (AIST) in Japan, we now have access to single-cell transcriptomics data of human immune cells that include the measurement of 29 lectins / glycan features (e.g., https://pubmed.ncbi.nlm.nih.gov/38164999/), compared to the prior state-of-the-art of one lectin. This project would extend our analyses to these much richer datasets, using a similar deep learning strategy as in our prior work, and also explicitly analyze the co-occurrence of lectin-binding on individual cells. Next to gaining much more insight into glycan functions in human immune cells, we ultimately also envision to use this information to partly restore information about the single-cell glycome from the measured glycan features on single cells, given corresponding bulk glycome measurements of the same samples. Research in this direction would thus pave the way for single-cell glycomics and functional glycomics studies on the single-cell level.

Physical lab attendance during the summer is not required, but we of course always welcome students, if there is interest.

*[Remote supervision possible]*

**2. Human X chromosome in disease – genetics and epigenetics**

**Supervisor:** Dr Tom Moore, School of Biochemistry and Cell Biology, [t.moore@ucc.ie](mailto:t.moore@ucc.ie)

The human X chromosome is implicated in diseases including reproductive disorders, intellectual disability and sexual differentiation. However, the X has been relatively neglected historically in linkage and association studies of disease because the different sex chromosome complements of males (XY) and females (XX) complicate the analysis. The recent T2T consortium sequencing of a full-length X chromosome provides a basis for further analyses of X chromosome structure-function relationships and disease associations. Genes near the X centromere are involved in intellectual disability and sexual differentiation may underpin sex differences in the incidence of neuropsychiatric conditions such as autism. This project aims to analyse X chromosome structure and gene expression with a particular focus on exploring possible links between autism and gender. Of particular interest is the possible role of transmissible epigenetic changes.

The project will use publicly available data resources and tools such as the Allen Brain Atlas, NCBI GEO, Ensembl, UCSC genome browser, to explore gene expression and regulatory relationships between autism and sexual differentiation. Similarly, publicly available genomics resources will be analysed to generate maps of structural variants in candidate gene regions, and to examine syntenic relationships with other species. There will be an initial specific focus on the androgen receptor gene region, which underpins a variety of gender-related diseases and alternative phenotypes, and which is genetically linked to multiple genes implicated in intellectual disability.

**3. Development of microbial metabolism in the first year of life**

**Supervisors:** Prof Liam O’Mahony & Hannah Devotta, School of Microbiology, [liam.omahony@ucc.ie](mailto:liam.omahony@ucc.ie)

The composition and metabolic activity of the early life microbiota contributes significantly to education of the immune system and protects against inappropriate immune reactivity to non-dangerous antigens, such as food proteins. Multiple studies have associated changes in microbiota composition with an altered risk of allergy. However, the metabolic pathways associated with these differences in microbiota taxa have not been well described. We have recently completed metagenomic sequencing of 350 infants at two timepoints (6 months and 12 months of age). In this project, we wish to map out the microbial metabolic networks that associate with microbiome maturation and identify those pathways that are different in infants with allergies compared to those that do not develop allergies. Our hypothesis is that specific genes encoding for immunomodulatory metabolites (e.g. genes required for tryptophan metabolism) will be present at different levels in infants with allergic diseases. This analysis will help us identify novel metabolites for future targeted interventions to prevent allergy.

*[Remote supervision possible]*

**4. Analysis of Uro-pathogenic *E. coli* Genome Sequences:**

**Supervisor:** Prof Michael Prentice, School of Microbiology, [m.prentice@ucc.ie](mailto:m.prentice@ucc.ie)

This project involves the phylogenetic assignment of 35+ *E. coli* genome sequences (sample collection still in progress at the time of writing) fromstrains isolated from infected urine. Genome sequences will be provided. As well as cgMLST typing, assignment of strains to phylogenetic groups will be required. Genome comparison with a previously published set of 47 *E. coli* sequences from my laboratory (PUBMED ID 31138611) is envisaged, with specific attention to phylogeny, the ethanolamine utilisation (*eut*) operon, other metabolic operons, and predicted antimicrobial resistances using AMR tools such as abricate, resfinder or RGI. There may be RNA-Seq analysis from laboratory experiments with some isolates. No wet laboratory work is required.

*[Remote supervision possible]*

**5. Mimicry: Exploring Novel Strategies in the Coevolutionary Arms Race between Hosts and Viruses**

**Supervisor:** Dr Máire Ní Leathlobhair, Computational Genomics Group, Trinity College Dublin, Máire Ní Leathlobhair, [nleathlm@tcd.ie](mailto:nleathlm@tcd.ie)

Viruses have evolved numerous methods, including host factor mimicry, to counter responses by host immune systems and to manipulate their hosts[1](https://paperpile.com/c/jJEDMs/CJwi). Mimicry of host factors gives viruses a better chance to survive and replicate in a host and spread within a population and is an important aspect of the ongoing "arms race" between viruses and their host organisms[2](https://paperpile.com/c/jJEDMs/E88u). Until recently, instances of viral mimicry were limited to mimics of immunomodulatory proteins and growth factors, but it has now been shown that viruses possess ahve the potential to encode at least 16 different peptides with high sequence similarity to several peptide hormones, including insulin[3,4](https://paperpile.com/c/jJEDMs/3C0B+MZwu). Notably, these findings underscore the vast potential for viruses to mimic a substantial proportion of regulatory peptides, as current research has predominantly focused on human peptide sequences.

The primary aim of this project is to leverage bioinformatic methodologies to uncover novel viral mimics. By doing this we can better understand mechanisms employed by viruses, elucidating biologically active mimetics that contribute to infection and disease. We will expand existing methodologies for identifying viral mimics to encompass sequence homology analysis alongside structural/functional homology[5,6](https://paperpile.com/c/jJEDMs/XPdX+iCEv). Using this comprehensive pipeline, we will explore the existence of viral mimics a broad range of important hormones. Furthermore, by investigating the activity and interactions of hormone mimics across various model organisms including mouse (*Mus musculus*), fruit fly (*Drosophila melanogaster*) and worm (*Caenorhabditis elegans*), we anticipate uncovering insights into evolutionary conservation and physiological adaptations.

References

1. [Alcami, A. Viral mimicry of cytokines, chemokines and their receptors. *Nat. Rev. Immunol.* 3, 36–50 (2003).](http://paperpile.com/b/jJEDMs/CJwi)

2. [Elde, N. C. & Malik, H. S. The evolutionary conundrum of pathogen mimicry. *Nat. Rev. Microbiol.* 7, 787–797 (2009).](http://paperpile.com/b/jJEDMs/E88u)

3. [Huang, Q., Kahn, C. R. & Altindis, E. Viral Hormones: Expanding Dimensions in Endocrinology. *Endocrinology* 160, 2165–2179 (2019).](http://paperpile.com/b/jJEDMs/3C0B)

4. [Altindis, E. *et al.* Viral insulin-like peptides activate human insulin and IGF-1 receptor signaling: A paradigm shift for host–microbe interactions. *Proceedings of the National Academy of Sciences* 115, 2461–2466 (2018).](http://paperpile.com/b/jJEDMs/MZwu)

5. [Drayman, N. *et al.* Pathogens use structural mimicry of native host ligands as a mechanism for host receptor engagement. *Cell Host Microbe* 14, 63–73 (2013).](http://paperpile.com/b/jJEDMs/XPdX)

6. [Cohen, L. J. *et al.* Corrigendum: Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* 556, 135 (2018)](http://paperpile.com/b/jJEDMs/iCEv).

*[Remote supervision possible]*

**6. Strain-level and metagenome assembled genome (MAG) analysis of bifidobacteria in the infant gut microbiota**

**Supervisors:** Dr Cathy Lordan, Profs Paul Cotter & Douwe van Sinderen, [Cathy.Lordan@teagasc.ie](mailto:Cathy.Lordan@teagasc.ie)

The proposed project focuses on bifidobacteria and the early life microbiota. Particular bifidobacteria species, such as *Bifidobacterium bifidum*, *Bifidobacterium breve, Bifidobacterium longum* subsp. *infantis* and *Bifidobacterium longum* subsp. *longum*, are some of the earliest colonisers of the early life gut microbiota. They play an integral role in interacting with the host immune system and developing the gut microbiota at this critical stage of life. In this project the student will conduct strain-level analysis, using tools such as StrainPhlAn or inStrain, on a shotgun metagenomic dataset comprising of 84 infant faecal samples. These samples were collected from infants over a period of 8 weeks (age 1 week to 8 weeks). They will also generate metagenome assembled genomes (MAGs) from this dataset in an assembly-based approach. High quality MAGs will be taxonomically classified and bifidobacteria members will be selected for additional analysis. This will give insights into the microbial composition and diversity which will complement the metagenomic analysis already completed. There is also opportunity to conduct functional analyses on these MAGs to determine the human milk oligosaccharide (HMO) related carbohydrate active enzymes (CAZymes) present using tools like dbCAN3 and eggNOG-mapper. This would allow a greater understanding of the potential oligosaccharide degradation abilities present in these infant-derived bifidobacteria MAGs.

**7. High-throughput image analysis to elucidate the effect of a polyQ expansion mutation in amyotrophic lateral sclerosis**

**Supervisor:** Dr Katja Burk, School of Biochemistry and Cell Biology, [KBurk@ucc.ie](mailto:KBurk@ucc.ie)

Work in our lab is focused on the neurodegenerative disease amyotrophic lateral sclerosis (ALS). Our work is stem cell based and relies heavily on high-resolution imaging data. To analyse the data fast, accurately, and reproducibly, we are looking for a student that can produce analysis scripts in Python, MatLab, or ImageJ. Wet lab work is not required since the data is already compiled.    
ALS is a neurodegenerative disease affecting motoneurons, leading to muscle weakness, bulbar dysfunction, and is ultimately fatal 3-5 years after disease onset. Our understanding of the genetic background of the disease has significantly improved, but inter-patient variability in age of onset, severity or penetrance cannot yet be fully described. Therefore, over the past years more focus has been put onto disease-modifying and risk factors – one of which is a poly-glutamine repeat expansion in the protein Ataxin-2. The protein is ubiquitously expressed and involved in RNA metabolism, stress response, and cellular homeostasis. Our data suggests that its functions converge on the cytoskeleton and could alter intracellular transport processes. We have acquired images that show the localization of cytoskeletal proteins, Ataxin-2, and potential interactors in induced motoneurons. Analysis should revolve around image segmentation, proximity of structures to each other, counts, intensities, co-localisations, etc. We also have live imaging data in which we are looking at intracellular trafficking events. In these, parameters like cargo speed and travelled distance should be analysed from kymographs. Presence is not required once biological parameters for analysis have been determined.

*[Remote supervision possible]*

**8. Beyond the Helix: Error Analysis and Mitigation Strategies for DNA-Based Data Storage**

**Supervisor:** Dr Md Noor-A-Rahim, School of Computer Science & IT, [md.noorarahim@ucc.ie](mailto:md.noorarahim@ucc.ie)

The exponential growth of digital data has spurred scientific exploration into alternative technologies for information storage. Recognized for its innate high density and long-lasting preservation capabilities, DNA has emerged as a promising medium for storing vast amounts of data. Despite its potential as a disruptive digital data storage solution, DNA storage is susceptible to errors, including insertions, deletions, and substitutions, which may occur during the synthesis and sequencing processes.

This project's primary objective is to comprehensively analyze errors in different types of DNA data storage. Additionally, the project seeks to develop techniques to mitigate the impact of errors in DNA storage, paving the way for more reliable and robust utilization of this innovative data storage medium.

*[Remote supervision possible]*

**9. Predict and understand the antimicrobial properties of thiourea organometallic complexes using molecular docking simulations**

**Supervisor:** Dr Davide Tiana, School of Chemistry, davide.tiana@ucc.ie

The World Health Organization (WHO) lists antimicrobial resistance (AMR) among top 10 threats for global health. In 2019 AMR causes almost 5 milion death. It has been estimated that, by 2050, AMR will cause10 million death by 2050, exceeding deaths caused by cancer. For this reason, there is a pressing importance to explore alternative antimicrobial drugs/ treatments. In this case, thiourea derivatives can play a significant role having shown promising antimicrobial properties. These molecules are of interest because of their versatility and capability to form organometallic complexes which can show enhanced antimicrobial activity.

In my research group we build libraries of thiourea complexes using a systematic analysis in which the substituents on the ligand are changed systematically and combined with different metals. We then use Molecular Docking to study the antimicrobial properties of these molecules. The purpose of using Molecular Docking is to predict the most likely ‘binding scenarios’ between a protein and a ligand. This way we can understand the mechanism that is behind the antimicrobial activity of molecules. After having elucidated the protein molecule interaction, we can design new molecules by functionalising of modifying its composition.

Objective: predict and/or explain the antimicrobial activity of existing and new computationally designed thiourea organometallic complexes using molecular docking.

Deliverables: rationalisation of the structure antimicrobial active property relationship of thiourea molecules.

### **10. Exploring the emerging paradigm of partitioned decision making in the nosocomial pathogen *Pseudomonas aeruginosa***

**Supervisor:** Dr Jerry Reen, School of Microbiology

LysR transcriptional regulators (LTTRs) are the largest family of regulatory protein encoded in species of microbial origin. Responding to host and endogenous signals as well as environmental cues, these proteins govern the virulence and biosynthetic potential of microorganisms, as well as acting as valuable biosensors for drug discovery. This research programme will advance ongoing research on the evolution of LTTRs and provide new insights into the biotechnological potential of these key regulatory systems.

The research project builds on the very recent finding that, despite the normal frequency of LTTR distribution across genomes, a large tract of the genome of the significant human pathogen *Pseudomonas aeruginosa*, has no LTTR protein encoded. This is the first time this has been reported (now published in Microbial Genomics with two past MSc Bioinformatics students included as authors) and it points a partitioning of decision making in the genomes of microbial pathogens. We now want to explore the extent to which this partitioning extends to other transcription factor families (GO analysis would suggest that it does) and to the genomes of other species/genera/phyla. The functional significance of this partitioning will also be explored through network analysis and metabolic modelling.

Key Milestones:

* Partitioning of LysR transcriptional regulation in *Pseudomonas aeruginosa* genomes
* Partitioning of other families of transcriptional regulation in *Pseudomonas aeruginosa* genomes
* Cross-species/genera analysis of transcriptional regulation distribution
* Systems based approach to understanding functional significance of partitioning

Key technologies:

* R
* MEGA Phylogenetics
* Comparative genomics
* Metagenomic mining
* Hierarchical cluster and network analysis

References

Deery, J., Carmody, M., Flavin, R., Tomanek, M., O’Keeffe, M., McGlacken, G. and F.J. Reen. Comparative genomics reveals distinct diversification patterns among LysR-type transcriptional regulators in the ESKAPE pathogen *Pseudomonas aeruginosa*. *Microbial Genomics* (In Press).

Reen, F. J., Barret, M., Fargier, E., O’Muinneacháin, M. and F. O’Gara. (2013). Molecular Evolution of LysR-type transcriptional regulation in *Pseudomonas aeruginosa*. *Molecular Phylogenetics and Evolution.* 66(3):1041-9.

Reen, F.J., Haynes, J.M., Mooij, M.J., and O’Gara, F. (2013). A novel LysR-type transcriptional regulator PA2206 is required for an effective oxidative stress response in the important nosocomial pathogen *P. aeruginosa*. *PLoS ONE* 8(1):e54479.

**11. Investigation of translation dynamics across the cell cycle via the re-analysis of single cell ribosome profiling data**

**Supervisor:** Prof Pasha Baranov & Mr Jack Tierney, School of Biochemistry and Cell Biology, [P.Baranov@ucc.ie](mailto:P.Baranov@ucc.ie)

Single-cell sequencing technologies have revolutionised our understanding of cellular

heterogeneity and gene expression dynamics within complex biological systems. Among these

methodologies, single-cell Ribosome Profiling (Ribo-Seq) stands out as a powerful tool to

decipher translation dynamics at the single-cell level. In this project which is proposed in

collaboration with EIRNA Bio, we aim to leverage publicly available (1) and in-house single-cell Ribo-Seq datasets to investigate translation dynamics throughout the cell cycle.

By analysing scRibo-seq data across different stages of the cell cycle, we seek to uncover dynamic changes in translational regulation that accompany cell cycle progression specifically focusing on translation termination. This investigation holds immense potential to unravel novel insights into the coordination between transcription and translation, shedding light on the

intricacies of cellular behaviour at a single-cell resolution.

Through this project, we offer an exciting opportunity for an incoming Master's student to delve

into cutting-edge scRibo-Seq methodologies and contribute to advancing our understanding of

translation dynamics in the context of cell cycle progression. The student will gain valuable experience in bioinformatic analysis, pipeline development, data integration, and interpretation,

thereby fostering skills essential for a career at the intersection of computational biology and

cellular research. This project will be carried out as a collaboration between LAPTI (https://lapti.ucc.ie) and EIRNA Bio (https://eirnabio.com/). The successful completion of the project may open future career opportunities either in academic or industrial environments.

Expectations from the student: The ideal student for this project would be eager to gain experience with a wide range of tools, be proficient in writing computer scripts (e.g. Python, R),

eager to learn how to analyse high-throughput sequencing data and ideally have basic understanding of eukaryotic translation. Physical attendance in the School of Biochemistry and

Cell Biology will be required for the successful completion of this project.

References

Van Insberghe et al., (2021) Single-cell Ribo-seq reveals cell cycle-dependent translational pausing

https://www.nature.com/articles/s41586-021-03887-4

Ingolia et al., (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using

ribosome profiling. <https://www.science.org/doi/10.1126/science.1168978>

**12. Exploiting Genome Scale Metabolic Models to design reengineered yeast for biotechnology**

**Supervisor**: Prof John Morrissey, School of Microbiology, [J.Morrissey@ucc.ie](mailto:J.Morrissey@ucc.ie)

Genome scale metabolic models describe the entire metabolism of an organism and are a useful tool to compare species and design engineering strategies to redirect metabolism for production of molecules of biotechnological interest. We are interested in the yeast *Kluyveromyces marxianus* as a chassis for industrial biotechnology and have developed a toolkit for engineering this yeast (Rajkumar et al., 2020; 2022). We previously collaborated in a EU-funded project that developed a new protein-constraint GEM for *K. marxianus* and other yeasts that is superior to existing models (Domenzain et al. 2022). We would like to now use this model to (i) compare the metabolism of *K. marxianus* with other biotechnological yeasts with particular emphasis on critical metabolic nodes; (ii) identify possible engineering strategies to optimise or reprogramme metabolism for production of specific metabolites that are of interest for ongoing and future projects.

References

Domenzain I, Sánchez B, Anton M, Kerkhoven EJ, Millán-Oropeza A, Henry C, Siewers V, Morrissey JP, Sonnenschein N, Nielsen J. 2022. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0.Nat Commun. 13(1):3766. doi: 10.1038/s41467-022-31421-1.

Rajkumar AS, Morrissey JP. 2022. Protocols for marker-free gene knock-out and knock-down in *Kluyveromyces marxianus* using CRISPR/Cas9. FEMS Yeast Res. 22(1):foab067. doi: 10.1093/femsyr/foab067

Rajkumar AS, Morrissey JP. 2020. Rational engineering of *Kluyveromyces marxianus* to create a chassis for the production of aromatic products. Microb Cell Fact. 19(1):207. doi: 10.1186/s12934-020-01461-7.

*[Remote supervision possible]*

**13. Genomic characterisation of novel intestinal *Blautia* bacteria**

**Supervisor:** Dr Hilary Browne, School of Microbiology, [hb4@sanger.ac.uk](mailto:hb4@sanger.ac.uk)

Humans are colonised by hundreds of different bacterial species with beneficial properties. Despite their importance, we still have a poor understanding of the metabolic capabilities and evolutionary history of most species in the intestinal microbiota. The massive increase in genomes from cultured isolates and MAGs (metagenome-assembled genomes), assembled directly from metagenomes now permits detailed analysis of individual species at strain-level resolution [1-3].

*Blautia* is a prevalent bacterial genus within the intestinal microbiota. Recent studies have linked *Blautia* specieswith diverse health benefits including pathogen inhibition, cognitive development and metabolic homeostasis [4, 5]. Despite this, the phylogenetic structure and functions of most *Blautia* species remain poorly understood. This project seeks to better characterise novel *Blautia* species by generating high resolution phylogenies and identifying encoded functional capabilities. Analysis will include: (1) screening genomes of selected novel *Blautia* species to remove low-quality genomes that are incomplete or contaminated and that could limit downstream analysis, (2) building detailed core gene phylogenies to understand, at strain-level, phylogenetic structure and evolutionary relationships between different species and, (3) functional annotation of genomes to identify metabolic pathways, antibiotic resistance genes and phenotypes that promote transmission such as sporulation [6].

This project will be physically based at UCC and it will not be possible to carry out remotely. The project has the potential to develop into a PhD.

References

1. Browne, H.P., et al., *Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation.* Nature, 2016. 533: p. 543.

2. Forster, S.C., et al., *A human gut bacterial genome and culture collection for improved metagenomic analyses.* Nature Biotechnology, 2019. 37(2): p. 186-192.

3. Almeida, A., et al., *A new genomic blueprint of the human gut microbiota.* Nature, 2019. 568(7753): p. 499-504.

4. Liu, X., et al., *Blautia—a new functional genus with potential probiotic properties?* Gut Microbes, 2021. 13(1): p. 1-21.

5. Caballero, S., et al., *Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium.* Cell Host & Microbe, 2017. 21(5): p. 592-602.e4.

6. Browne, H.P., et al., *Transmission of the gut microbiota: spreading of health.* Nature Reviews Microbiology, 2017. 15: p. 531.

**14. Spatial patterning in barnacles**

**Supervisor:** Dr Markus Eichhorn, School of Biological, Earth and Environmental Sciences, [markus.eichhorn@ucc.ie](mailto:markus.eichhorn@ucc.ie)

Barnacles are a common feature of intertidal habitats, colonising partially-submerged rock surfaces. The spatial patterns they form are generated by a combination of processes including their interactions at establishment followed by growth and competition for space. There are multiple species that co-occur on the coastline of Cork, each of which has different scales and intensities of interactions. They are therefore an excellent system for studying the emergent properties of self-assembling ecological communities. We are seeking a method for identifying individual barnacles from standardised photographs and measuring key parameters such as the size and orientation of their opercular plates, distinguishing among species, and recognising those that are dead. The effectiveness of any automated approach can be tested against existing manually-collected data. The expected output would be a script that can efficiently and accurately process large numbers of photographs. A potential starting framework is the ML-morph package which is implemented in Python (<https://github.com/agporto/ml-morph>; Porto & Voje 2020).

Ref <https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13373>

*[Remote supervision possible]*

**15. Visceral Adiposity in GDM as a Mediator of Future Cardiometabolic Risk**

**Supervisor:** Dr Cathal McCarthy, Pharmacology & Therapeutics, [cmccarthy@ucc.ie](mailto:cmccarthy@ucc.ie)

Gestational diabetes mellitus (GDM) refers to abnormal glucose tolerance with onset during pregnancy and is the most common medical complication of pregnancy. The global prevalence of GDM continues to rise due to increases in the rates of obesity in women of reproductive age. GDM is also a major risk factor for future Type II diabetes and cardiometabolic disease in both mother and child. Visceral adipose (fat) tissue (VAT) which surrounds our internal abdominal organs is an important site for lipid storage and is recognised as metabolically active yet only accounts for 6-20% of total fat mas. Increased visceral adiposity is often evident in obesity and is also an early indicator of metabolic disturbance and insulin resistance, as evident in Type II diabetes. Given the importance of visceral adipose tissue in the pathology of obesity and Type II diabetes, we are also keen to investigate its role in the development of GDM. As part of our study, we recruited women with a healthy uncomplicated pregnancy (classified as normal glucose tolerant NGT) and GDM participants, and categorised them according to their BMI, into either nonobese (< 30kg/m2) or obese (≥ 30kg/m2) cohorts. We have completed RNA seq of visceral adipose tissue from women in both cohorts and have identified significant differences in the expression of certain gene signatures and associated pathways.

Aim of Research Project: Given the biological link between GDM and Type II diabetes, in this research project, we would like to compare our RNA-seq data with publicly available RNA seq data from Type 2 diabetes and/or obesity studies to see if there are any common associated signatures and determine which of these signatures are mediated by an elevated BMI.

*[Remote supervision possible]*

**16. Could “obelisk” viroids play a role in inflammatory bowel disease – a novel metatrancriptomic exploration of a brand new molecule**

**Supervisor:** Prof Marcus Claesson, School of Microbiology, [m.claesson@ucc.ie](mailto:m.claesson@ucc.ie)

This year researchers at Stanford discovered a completely new, and potentially revolutionary, type of molecule in the human microbiome. So-called “obelisks” are a previously unknown class of viroids, which are roughly 1000 bases of circular RNA that form rod-shaped secondary structures, coding for a novel protein superfamily. As obelisks contain genes that are unlike any discovered so far in other organisms, they comprise a class of diverse RNAs that have colonized, and gone unnoticed in, human, and global microbiomes. The authors recently found these obelisks in 7% and 50% of human gut and oral metatranscriptomes, respectively (1) (2) (3).

So far, it is unknown how obelisks affect human health. The student selecting this project will therefore explore both unpublished metatranscriptome data from 27 colonic biopsies from the same samples as in 2), as well as published data from stool samples (HMP2-IBD; 3). The bioinformatic approach developed by 1) will be followed in order to associate disease states to the presence and type of obelisks discovered. Indicative findings can be followed up in grant applications to facilitate new targeted or metatranscriptomic studies.

References

1. Zheludev IN, Edgar RC, Lopez-Galiano MJ, de la Peña M, Babaian A, Bhatt AS, Fire AZ. Viroid-like colonists of human microbiomes. bioRxiv [Preprint]. 2024 Jan 21:2024.01.20.576352. doi: 10.1101/2024.01.20.576352.

2. ‘It’s insane’: New viruslike entities found in human gut microbes. Science. Jan 2024. doi: 10.1126/science.znxt3dk

3. Sidik S. 'Wildly weird' RNA bits discovered infesting the microbes in our guts. Nature. 2024 Jan 29. doi: 10.1038/d41586-024-00266-7

4. Ryan, F.J., Ahern, A.M., Fitzgerald, R.S. et al. Colonic microbiota is associated with inflammation and host epigenomic alterations in inflammatory bowel disease. Nat Commun 11, 1512 (2020). https://doi.org/10.1038/s41467-020-15342-5

5. Lloyd-Price J, et al, Xavier RJ, Huttenhower C. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019 May;569(7758):655-662. doi: 10.1038/s41586-019-1237-9.

**17. Computational Analysis of Bacterial Strains isolated from faecal samples collected in rural Papua New Guinea: Insights into their Evolution and Probiotic Potential**

**Supervisors:** Prof Jens Walter & Dr John Leech, School of Microbiology, [jenswalter@ucc.ie](mailto:jenswalter@ucc.ie)

Papua New Guinean individuals, like other populations with non-Westernized lifestyles, possess more diverse microbiomes and exhibit a lower incidence of chronic diseases. Consequently, their microbial isolates hold particular interest for their potential health-promoting properties. We have isolated bacterial strains from faecal samples collected in rural Papua New Guinea. Several of these strains belong to species for which there is evidence that they are depleted in industrialized microbiomes (so called VANISH species). These genomes provide a unique opportunity to answer some interesting question on the evolution and ecology of VANISH species. In addition, there is potential to develop these strains into probiotics.

This study presents a comprehensive bioinformatics investigation aimed at characterizing microbial strains isolated from individuals in Papua New Guinea. Leveraging computational approaches for genome assembly and annotation, the project focuses on elucidating the phylogeny and genetic landscape of these isolates with focus on the evolution and ecology of these species. Particular emphasis will be placed on genes associated with probiotic properties and carbohydrate-active enzymes (CAZymes). The investigation aims to shed light on the strains' potential probiotic functionalities and their ability to metabolize diverse carbohydrate substrates. The genomic data analysis involves de novo assembly of sequenced genomes followed by phylogenomic analysis and a thorough annotation to identify genes of interest, including those encoding probiotic factors and CAZymes. Furthermore, the study explores additional genes relevant to microbial physiology and potential applications in biotechnology. Through systematic analysis, this project seeks to uncover the genetic basis underlying the strains' potential probiotic properties and their capacity to utilize various carbohydrate sources. Understanding these genetic determinants could provide valuable insights into the strains' potential applications in promoting human health and in biotechnological processes.

This bioinformatics investigation serves as a crucial component of a larger project, where insights gleaned from computational analyses will be subsequently validated through experimental fermentation studies using a BioLector, a high-throughput fermentation system. Integration of computational and experimental approaches offers a comprehensive understanding of the Papua New Guinean isolates' characteristics and potential applications, thus contributing to advancements in both bioinformatics and biotechnology.

**18. Catching the Silent Killer: Investigation of Non-Coding RNAs in the Single Cell Landscape of Ovarian Cancer**

**Supervisor:** Ms Aideen McCabe & Dr Kellie Dean, School of Biochemistry and Cell Biology, [K.Dean@ucc.ie](mailto:K.Dean@ucc.ie)

Ovarian cancer is the most fatal gynaecological malignancy, accounting for over 200,000 deaths worldwide each year (Sung et al., 2021). Most cases of ovarian cancer are diagnosed at a late stage, mainly due to non-specific symptoms and a lack of specific and sensitive early detection markers. Current screening approaches for ovarian cancer involve serum cancer antigen 125 (CA125) measurement, however this method has shown limited effectiveness in reducing mortality (Trinidad et al., 2020). CA125 is also a non-specific biomarker, often increased in various benign conditions (Sevinc et al., 2007). Therefore, there is a pressing need for early detection biomarkers in ovarian cancer to enhance diagnostic accuracy and reduce diagnostic delay.

Recent advancements in single-cell RNA sequencing (scRNA-seq) technology have revolutionized our understanding of intratumoral heterogeneity and molecular dynamics in cancer. Additionally, non-coding RNAs (ncRNAs) have been highlighted as promising biomarkers in cancer due to their relative stability in circulation, tissue specificity (Badowki et al., 2022). In this project we propose an *in silico* analysis of non-coding RNAs in the single cell landscape of ovarian cancer.

Publicly available scRNA-seq datasets will be identified and downloaded from the Gene Expression Omnibus and The European Nucleotide Archive. Leveraging the high resolution and sensitivity of single cell transcriptomics, this data will be used to identify if known ncRNA biomarkers exhibit differential expression patterns or cellular localization within tumours. An additional aim of this project is to and identify novel markers with diagnostic, prognostic, and therapeutic implications in ovarian cancer. The student will gain experience using Seurat, a popular R package for analysis of single-cell data, learning skills such as quality control of data, normalization, scaling, dimensionality reduction using PCA, non-linear dimension reduction for visualization, clustering of cells and annotation of clusters. The overall goal of this study is to explore the distribution of non-coding RNAs at the single cell level, with the aim of identifying and validating early diagnostic biomarkers.

References

Badowski, C., He, B. and Garmire, L.X., 2022. Blood-derived lncRNAs as biomarkers for cancer diagnosis: the Good, the Bad and the Beauty. NPJ precision oncology, 6(1), p.40.

Sevinc, A., Adli, M., Kalender, M.E. and Camci, C., 2007. Benign causes of increased serum CA-125 concentration. The lancet oncology, 8(12), pp.1054-1055.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 71(3), pp.209-249.

Trinidad, C.V., Tetlow, A.L., Bantis, L.E. and Godwin, A.K., 2020. Reducing ovarian cancer mortality through early detection: approaches using circulating biomarkers. Cancer Prevention Research, 13(3), pp.241-252.

#### **19. In-silico design of circular RNA therapeutics for modulation of gene and protein expression in sepsis**

**Supervisor:** Dr Piotr Kowalski School of Pharmacy, [piotr.kowalski@ucc.ie](mailto:piotr.kowalski@ucc.ie)

Sepsis is a dysregulated host response to infection that can lead to life-threatening multiple-organ failure. Annually, sepsis affects about 50 million people of all ages worldwide, resulting in an estimated 11 million deaths. Yet, therapeutic options for this condition are limited due to the complex and systemic nature of the disease and the lack of effective pharmacological interventions. My group aims to develop an effective therapeutic intervention for the treatment of sepsis-induced multiple organ failure based on the delivery of Circular RNAs (circRNAs) which are a new class of non-coding RNAs with a unique closed-loop structure that could help address the current limitations of the RNA drugs in the disease context and open new therapeutic avenues (1). Therapeutic delivery of engineered synthetic circRNAs can allow taking full advantage of their unique features and functions, including increased intracellular stability, the ability to affect multiple biological pathways by sponging microRNA, and their potential for cellular context-specific control of protein expression via internal ribosome entry site (IRES)-mediated cap-independent translation (2).

To better guide the design of therapeutic circRNA for tissue-specific protein translation we synthesized a library circRNA containing 40 unique IRES sequences of viral and mammalian origin that are evaluated for their ability to preferentially drive potent expression in diseased endothelial cells (ECs). The student will utilise this data and computational tools (3) (e.g. RNA fold, RPBmap) to investigate IRES structure and study RNA-protein interactions with IRES-Trans-Acting Factors (ITAFs) which can assist cap-independent translation, mapping protein binding sites on selected IRESs. In-silico findings will be subsequently validated by identification of proteins specifically bound to a given circRNA, using a method combining RNA pull-down with SILAC high-throughput mass spectrometry (RP–SMS).  
Inspired by endogenous circRNA sponges, we are also interested in developing synthetic circRNA miRNA sponges engineered to specifically modulate the biological pathways, which can be an attractive approach for targeting complex diseases like sepsis where entire networks of genes are dysregulated. To guide target selection for circRNA miR sponge-based intervention in sepsis, students will identify potential microvascular miRNA targets by exploring the miRNome dataset, published by our clinical collaborators, (4) and publicly available RNA-seq databases, including the human microRNA disease database, miRBase, and miRmine.

References

1. Young TL, Wang KC, Varley AJ, Li B, Clinical delivery of circular RNA: Lessons learned from RNA drug development. Adv Drug Deliv Rev. (2023).
2. Wesselhoeft, R.A., Kowalski, P.S. and Anderson, D.G, Engineering circular RNA for potent and stable translation in eukaryotic cells. Nature Communications (2018).
3. Sneha Bheemireddy, Sankaran Sandhya, Srinivasan, N. and Ramanathan Sowdhamini, Computational tools to study RNA-protein complexes. Frontiers in Molecular Biosciences (2022)
4. Luxen M, Zwiers PJ, Meester F, Jongman RM, Kuiper T, Moser J, Pultar M, Skalicky S, Diendorfer AB, Hackl M, van Meurs M, Molema G, Unique miRNome and transcriptome profiles underlie microvascular heterogeneity in mouse kidney. Am J Physiol Renal Physiol (2023)

**20. Predicting clinical relapse in inflammatory bowel disease patients**

**Supervisors:** Kiang Wei Kho & Salvatore Tedesco, Tyndall National Institute, [kiangwei.kho@tyndall.ie](mailto:kiangwei.kho@tyndall.ie)

There are about 10 million people worldwide living with inflammatory bowel disease (IBD). Currently, IBD remains incurable and is characterized by alternating periods of remission and relapse. As such, the aim of therapy is to induce remission for as long as possible. Unfortunately, disease flares can still occur randomly and are mostly unpredictable. Here, you will be part of our ongoing project at Tyndall, in which the spatial distribution of biomarkers (proteins or bacteria) in the mucosal biopsy samples from treatment-responsive patients, will be mapped out using luminescent nanoparticles. In this Msc project, you will perform quantitative analysis on the acquired images with an aim of identifying causal correlations between structural features, biomarker-expression level, and the time (within or longer than 6 months) to relapse since remission. Techniques that will be used include image segmentation, spatial filtering for background correction and feature enhancement, feature selection, causal forest. You should have good proficiency in MATLAB/Python.

**21. Discovery of novel transducing prophages in the human gut microbiome**

**Supervisor:** Andrey Shkoporov, School of Microbiology, [andrey.shkoporov@ucc.ie](mailto:andrey.shkoporov@ucc.ie)

Complex natural microbial communities, such as those found in the human microbiome, are thought to undergo evolutionary changes primarily through horizontal gene transfer (HGT). This process bypasses the constraints of vertical evolution and enables individual strains to access a shared gene pool known as the distributed pangenome. However, there is a significant gap in our understanding regarding the frequency, impact, and underlying mechanisms of HGT within the microbiome. Our initial findings suggest that among the three main HGT mechanisms—namely, transformation by naked DNA, conjugation, and bacteriophage-mediated transduction—the latter predominates both in terms of its relative contribution to HGT and its rate, exerting a substantial influence on the microbiome [1,2]. The phageome, comprising a vast array of virulent and temperate bacteriophages, constitutes a crucial component of the microbiome, constituting approximately 5% of total microbial DNA [3]. Recent evidence from our research indicates that substantial quantities of bacterial genomic DNA are associated with the physically isolated phageome fraction from the human gut. Importantly, we have demonstrated that this association is not merely a result of random contamination but rather signifies a selective enrichment of DNA from specific bacterial taxa and genomic regions.

The overall goal of the project:

This project aims to undertake the genome assembly, annotation, and taxonomic evaluation of various bacterial isolates obtained from the faecal samples of healthy donors. Additionally, we will investigate the phageome of these individuals to identify instances of prophage induction and host DNA packaging facilitated by bacteriophages. The project's overarching objective is to evaluate the significance of bacteriophages in mediating horizontal gene transfer events within the human gut.

Specific objectives:

1. Assemble genomes using sequence reads derived from bacterial isolates sourced from human faecal samples.
2. Annotate genes and conduct taxonomic assignments on the assembled bacterial genomes.
3. Predict prophage regions.
4. Align phageome sequence reads with bacterial genomes to identify prophage induction and transduction events within the human gut microbiome.

Outcome: This project's findings will produce preliminary data for Work Package 1 (WP1) of the PHAGENET project, which seeks to elucidate the encapsulation of bacterial DNA within the human gut phageome.

References

1. Lerner, A.; Matthias, T.; Aminov, R. Potential Effects of Horizontal Gene Exchange in the Human Gut. *Front Immunol* 2017, *8*, 298585, doi:10.3389/FIMMU.2017.01630/BIBTEX.

2. Borodovich, T.; Shkoporov, A.N.; Ross, R.P.; Hill, C. Phage-Mediated Horizontal Gene Transfer and Its Implications for the Human Gut Microbiome. *Gastroenterol Rep (Oxf)* 2022, *10*, doi:10.1093/GASTRO/GOAC012.

3. Shkoporov, A.N.; Hill, C. Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host Microbe* 2019, *25*, 195–209, doi:10.1016/J.CHOM.2019.01.017.

**22. Decoding Synergistic Cell Death Mechanism: RNA-Seq Analysis of CDK12 and BRD4 Inhibition in Triple-Negative Breast Cancer**

**Supervisor:** Malgorzata Krajewska, School of Biochemistry and Cell Biology, [malgorzata.krajewska@ucc.ie](mailto:malgorzata.krajewska@ucc.ie)

# Triple-negative breast cancer (TNBC) poses a significant challenge in oncology due to its limited therapeutic options. Thus, there is an urgent need for innovative treatment strategies to enhance outcomes in TNBC patients. Transcription-associated cyclin-dependent kinases (CDKs), such as CDK12, have long been recognized for their role in transcription regulation and as potential therapeutic targets in cancer. However, the development of resistance to single-agent targeted therapies is inevitable.

# To address this issue, we screened a library of inhibitors currently in clinical trials and identified novel drug combinations that exhibit enhanced cytotoxicity through the inhibition of CDK12 and bromodomain-containing protein 4 (BRD4), an epigenetic regulator. Our findings demonstrate that this combination induces greater cytotoxicity compared to single-agent treatments. To elucidate the underlying mechanisms, we conducted RNA sequencing on breast cancer cells treated with individual drugs and the combination, providing novel insights for further investigation.

# The proposed MSc project focuses on the integrative analysis of RNA sequencing data to uncover the molecular mechanisms driving enhanced cytotoxicity in TNBC cells treated with CDK12 and BRD4 inhibitors. This involves processing raw sequencing data, conducting differential gene expression analysis, exploring pathway enrichment using gene set enrichment analysis (GSEA) and protein-protein interaction networks (STRING), and employing the GeneWalk method to identify candidate genes. By examining the impact of CDK12 and BRD4 inhibition on oncogenic and immune signatures, the research aims to unveil the synergistic cell death mechanisms. Additionally, transcription factor enrichment analysis will be conducted to identify potential new regulators of oncogenic signatures (e.g. muMerge). Furthermore, the significance of key genes identified in combination treatment will be assessed across different breast cancer subtypes (e.g. cBioPortal, The Cancer Genome Atlas), and their correlation with patient survival (e.g. R2 database) will be analyzed, providing valuable insights for the development of novel anticancer therapies.

# This project offers an exciting opportunity for an MSc student to contribute to the advancement of precision medicine and potentially revolutionize treatment strategies for aggressive cancers like TNBC.

References

1. Krajewska, M. *et al.* CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat Commun* 10, 1757, doi:10.1038/s41467-019-09703-y (2019).
2. Bayles, I. *&* Krajewska M*. et al.* Ex vivo screen identifies CDK12 as a metastatic vulnerability in osteosarcoma. *J Clin Invest* 129, 4377-4392, doi:10.1172/JCI127718 (2019)
3. Gao Y\*, Volegova M\*, Nasholm N, Das S, Kwiatkowski N, Abraham BJ, Zhang T, Gray NS, Gustafson C, Krajewska M\*\* and George RE\*\*, Synergistic anti-tumor effect of combining selective CDK7 and BRD4 inhibition in neuroblastoma, *Frontiers in Oncology,* 2022.
4. Ietswaart, R et al. GeneWalk identifies relevant gene functions for a biological context using network representation learning. Genome Biol. (2021)
5. Rubin, D et al., Transcription factor enrichment analysis (TFEA) quantifies the activity of multiple transcription factors from a single experiment. Com. Biology, (2021)

# Zimmerli, D. et al., MYC promotes immune-suppression in triple-negative breast cancer via inhibition of interferon signaling. Nat Comm, doi: 10.1038/s41467-022-34000-6, (2022).

**23. Investigating genomic diversity and host colonisation features of Dolosigranulum pigrum, a commensal bacterium in the nasopharyngeal microbiome**

**Supervisor:** Dr Francesca Bottacini, Munster Technological University, [Francesca.Bottacini@mtu.ie](mailto:Francesca.Bottacini@mtu.ie)

The human nasopharyngeal microbiota represent a complex ecosystem situated at the upper part of the respiratory tract and plays a crucial role in maintaining respiratory health by contributing to immune function, pathogen defense, and nutrient metabolism. Understanding the composition and dynamics of the nasopharyngeal microbiome is essential for unraveling its role in health and disease for the human host.

Among the myriad of microbial inhabitants in the nasopharynx, *Dolosigranulum pigrum* has gained increasing attention due to its potential role in modulating host immune responses and protecting against respiratory infections. As a Gram-positive bacterium belonging to the Bacillota phylum, *D. pigrum* has been shown the potential of acting as a beneficial commensal, often coexisting with other members of the microbial community and inhibit pathogen colonisation in the respiratory tract. Despite its importance, the precise mechanisms by which this bacteriuminteracts within the nasopharyngeal environment still remain poorly understood.

This bioinformatic project aims at identifying colonization features and ecological niche interactions of *D. pigrum*. By availing of comparative pangenome analysis methods the project seeks to delineate the genomic landscape of *D. pigrum* across several publicly available strains, shedding light on its core and dispensable genome features. By elucidating conserved genetic elements within the core genome associated with colonization persistence and niche adaptation, as well as exploring genomic variations within the dispensable genome, the project aims at providing key insights into the molecular mechanisms driving the interactions of this bacterium within its ecological niche.

Overall this study will advance our understanding of host-microbe dynamics in the nasopharyngeal environment and the potential beneficial activity of this novel bacterium in the human respiratory tract.

*[Remote supervision possible]*

# **24. Analysis of a novel virulence regulon in large plasmids in two major clonal lineages of enterotoxigenic Escherichia coli (ETEC)**

**Supervisor:** Prof Åsa Sjöling & Kaisa Thorell, Department of Chemistry and Molecular Biology, University of Gothenburg, [asa.sjoling@gu.se](mailto:asa.sjoling@gu.se), [kaisa.thorell@gu.se](mailto:kaisa.thorell@gu.se)

Enterotoxigenic *Escherichia coli* (ETEC) is a water- and food-borne bacteria that causes mild to severe diarrhea. Despite recent efforts to develop vaccines and improve water quality it remains a major cause of child mortality in middle- and low-income countries. The bacteria cause diarrhea by adhering to the epithelium in the lower part of the small intestine and secrete toxins e.g heat stable toxin, STh and STp, and/or heat labile toxin, LT that both cause a massive outflux of water and electrolytes from epithelial cells by deregulation of internal cAMP and cGMP levels. ETEC can express a large variation of different colonization factors (CFs) but usually CFs are expressed in combinations of one to three defined CFs in individual strains. The CFs and toxins are present on large virulence plasmids that are conserved within clonal lineages (von Mentzer *et al.,* 2014; von Mentzer *et al*., 2021).

Two of these lineages, L1, that expresses the ETEC toxins LT and STh and the colonization factors CS1, CS3, and CS21, and L2 that expresses LT, STh, CS2 + CS3 are associated with diarrheal disease in children and adults and all strains that express these virulence profile appear to be clonal and derived from one single ancestral isolate that seemingly emerged approximately ca 100 years ago and then managed to spread globally. Although genetically similar, strains of L1 and L2 differ in plasmid composition. L1 strains have 4-5 plasmids while L2 have 3 plasmids. Both lineages share one plasmid encoding the toxins and CS1.

The project will include analysis of plasmids with a focus on the conserved plasmid in both L1 and L2 in a set of previously sequenced L1 and L2 isolates. The data is generated by PacBio sequencing, and the student will be provided with assembled and annotated genomes. The major aim is to study a newly identified putative virulence operon. The operon *shf-rfbU-virK-msbB2* has been described to be involved in membrane stress regulation, intracellular survival, and biofilm formation in other species (*Shigella, Salmonella* and other enterobacteriaceae) but has not been characterized in ETEC before. Some of the genes have homologs on the chromosome, eg *MsbB* (also called *LpxM*), which is involved in LPS modification, and *shf*, (also known as *icaB* or *pgaB*) that are also present on the chromosome and in other species. The operon seems to be regulated by presence of MgCl2 in *Shigella spp.*

More specifically, the tasks will include analysis of the operon in ETEC and comparison of the entire operon as well as individual genes within ETEC genomes as well as with other species by using local and global alignments and analyses of presence of similar genes eg BLAST, MAUVE or similar. Location of both plasmid and chromosomal variants of the genes needs to be determined both in the ETEC genomes provided and in other species. Phylogenetic analyses of the sequences will be performed to determine genetic relationship between species. Since regulation of the operon and/or individual genes is expected, a search for transcription factor sites, start and stop codons and Shine-Dalgarno sequence in the DNA sequences of the ETEC genomes should be performed and compared to eg *Shigella* spp. Visualization of the genes, promoters and entire plasmids should be performed.

The project aims to describe the genetic composition of this new operon. It requires some background knowledge in microbiology and an interest in prokaryotic gene regulation and genomics. The project will be supervised via zoom and with more frequent meetings early in the summer and in September, hence the student is expected to work independently in July.

*[Remote supervision possible]*

**25. Elucidating Adrenomedullin signalling pathways in intestinal epithelial cells and macrophages from publicly available RNASeq datasets from** **Crohn’s Disease patients**

**Supervisor:** Silvia Melgar, APC Microbiome Ireland, [S.Melgar@ucc.ie](mailto:S.Melgar@ucc.ie)

Background: Crohn’s Disease (CD) is a chronic inflammatory condition of the gastrointestinal (GI) tract. Although aetiology is unknown, the impact of gut microbiota and genetic, inflammatory, and environmental factors has already been described (Friedrich, Pohin, Powrie 2019). CD has become a global disease and, as it is a lifelong condition, it is a burden to both patients and public health. CD is clinically stratified by location with distinct differences in presentation and risk for progression (Dulai et al, 2019). Ileal CD compared to colonic CD poses a higher risk for delayed diagnosis and potential complications, and with its different immune response profile, it is an important target for new treatment strategies, especially as it has been shown for different biological therapies that there is trend of lower efficacy in ileal CD compared to colonic CD (Atreya and Siegmund, 2021).

Adrenomedullin (AM) is a small peptide hormone with various biological functions including vasodilatation, angiogenesis, anti-inflammation, and potentiation of host defences against microbes. AM is highly produced by cells of the GI tract, including enteroendocrine cells, and is regulated by several stimuli including inflammatory mediators (Martínez-Herrero and Martínez, 2022). Pilot studies in patients with ulcerative colitis and CD supplemented with AM reported a reduction in clinical symptoms, accelerated mucosal healing, and maintenance of clinical remission over time (Kita et al, 2021; 2022).

Considering that mechanisms of action of AM in the GI tract are not well known, the aim of this project is to study differences in expression levels of AM, its known receptors and other noncanonical receptors in datasets from CD patients. Gathering this information will help understand the role of intestinal epithelial cells and macrophages in AM signalling pathway, but also contribute to our hypothesis of intercellular communication between different cell types as AM’s mechanism of action in GI tract.

Objectives:

1. To identify the alterations in expression of adrenomedullin and its known receptors in human GI tract from patients with CD compared to controls by using publicly available data sets from whole tissue and single cells.
2. Compare expression levels in different locations of human GI tract (small intestine vs colon) with the focus on characteristics of ileum.
3. Elucidate potential signalling pathways related to adrenomedullin's effect on e.g. barrier function and healing.

**26. Identifying epithelial cell subsets and pathways altered in the intestine of subjects with obesity and diabetes**

**Supervisor:** Silvia Melgar, APC Microbiome Ireland, [S.Melgar@ucc.ie](mailto:S.Melgar@ucc.ie)

Background: Obesity is a chronic condition characterized by excess adiposity that may be accompanied by different structural and functional abnormalities and with increased comorbidity and premature mortality risk thus reducing quality of life (Jastreboff et al., 2019; Lustig et al., 2022). The main cause of obesity is the energetic imbalance due to increased caloric intake and little expenditure. This induces metabolic and hormonal changes e.g. increase in blood sugar levels that induces a prediabetic status with increased risk of developing type 2 diabetes mellitus (T2DM), heart disease, and stroke. T2DM is a major non-communicable disease and one of the world’s fastest growing health problems, with a projected increase in the number of diabetic patients to 700 million by 2045 (Saeedi et al., 2019). T2DM is associated with significant morbidity, including increased risk of cardiovascular diseases (CVD) and stroke, hypertension, etc. These place an enormous burden on individuals, society, and the healthcare system (Brorsson & Pociot, 2015). T2DM is a non-reversible but preventable condition with overweight and obesity being major risk factors. The onset of T2DM is gradual, with most individuals progressing from normoglycaemia through a pre-diabetic state.

There is substantial evidence for the role of gut microbiota and impaired barrier in metabolic diseases including T2DM (Brunkwall & Orho-Melander, 2017). Recent clinical trials using Glucagon-like peptide-1 (GLP-1) receptor agonists (an incretin hormone produced by the enteroendocrine L cells in the distal intestine) have shown benefits to patients with these conditions. However, a role for other intestinal epithelial cell subsets in obesity and diabetes is yet to be determined. Thus, the aim of this project is to 1) identify alterations in intestinal epithelial cell subsets, and 2) pathways contributing obesity and/or diabetes pathophysiology. For this purpose, the student will be performing a bioinformatics analysis on in house and publicly available bulk and single cells RNASeq data sets. Findings from this project, will provide a better understanding on how gut epithelial cells, which are in closed contact to the microbiota, might support and/or sustain the development of these chronic conditions.

**27. A snapshot of bacterial evolution: population biology of sympatric *Shewanella* spp. In the context of ecological specialization**

**Supervisor:** Jonatan Martin Rodriguez, Department of Microbiology, Tumour & Cell Biology, Karolinska Institute, [jonatan.martin.rodriguez@ki.se](mailto:jonatan.martin.rodriguez@ki.se)

Our team has recently described a novel concept in bacterial physiology termed “Respiration-Induced Biofilm Formation” (RIBF) [1, 2] within bacteria of the genus *Shewanella*. We hyphothesize that RIBF might direct subgroups of strains to preferentially occupy certain environmental niches through biofilm formation and specialization of metabolic functions related to alternative electron acceptor utilization. We aim to interrogate this phenomenon in a sympatric population of *Shewanella* spp. occupying distinct niches within the same stratified ecosystem.

We have retrieved approximately >75 co-occurring *Shewanella* isolates from Baltic Sea sediments. Superficial cores were collected and *Shewanella* spp. were isolated at different depths (0, 2, 4 and 6 cm). All isolates are whole genome sequenced (draft genomes). The chosen depths represent both oxygenated layers and likely anoxygenic strata, given the rapid depletion of oxygen within such a stratified environment. Moreover, environmental DNA samples were extracted from representative samples of each depth, in duplicate, and sequenced. Preliminary analyses suggest that strains within this population exhibit distinct metabolic capacities and high levels of horizontal recombination.

Project description:

The proposed project involves genomic, metagenomic and taxonomic studies to characterize crucial aspects of the ecophysiology, evolution, and adaptation of coexisting bacterial populations. Specifically, the following points will be addressed:

1) First, genome-based phylogenies will be inferred and the taxonomic positioning of the strains will be related to their ecological niche and relevant phenotypes studied in our lab.

2) Next, a pangenome analysis will be carried out for the *Shewanella* population using established pipelines (e.g. Panaroo ([3]). The core and accessory genomes will be defined. We foresee that each taxonomic genospecies (or genomovars) may have adaptive features that make it divergent from each other. Gene content will be categorized according to clusters of orthologous genes (COGs) or KEGG pathways to gain an overview of their biological functions.

3) Many bacterial species share genetic modules that provide ecological fitness through horizontal gene transfer. Such events are assumed to be more frequent between coexisting strains. Horizontal recombination events and recombination hotspots in whole-genome sequences will be investigated using state-of-the-art inference algorithms [4, 5]. The genomic regions that are more prone to recombination will be investigated in relation to the pangenome.

4) The degree of conservation of respiration pathways for ecologically relevant electron acceptors (e.g. oxygen respiration, nitrate/nitrite reduction, and respiration of N- and S-oxides) will be investigated through the analysis of the terminal reductases for these compounds, identified by BLAST-based searches or homology modeling [6]. We have previously documented selective pressures involving loss or acquisition of chromosomally encoded respiration capacities [2, 6], which are likely to occur here in relation to ecological specialization.

5) In this regard, *Shewanella* ecotypes will be delineated using established algorithms [8, 9] which predict ecologically distinct bacterial populations from genome sequence data.

6) Finally, metagenome sequence analysis will reveal the relative abundance of *Shewanella* spp. in the sedimental bacterial community, as well as the presence and abundance of genes related to respiration of distinct electron acceptors, in relation to those encoded in *Shewanella* genomes. This analysis will contextualize the *Shewanella* population with the broader microbial community at each studied depth and will give a notion of their ecological significance.

1. Martin-Rodriguez AJ: Respiration-induced biofilm formation as a driver for bacterial niche colonization. *Trends Microbiol* 2023, 31(2):120-134.

2. Martin-Rodriguez AJ, Reyes-Darias JA, Martin-Mora D, Gonzalez JM, Krell T, Romling U: Reduction of alternative electron acceptors drives biofilm formation in Shewanella algae. *NPJ Biofilms Microbiomes* 2021, 7(1):9.

3. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA *et al*: Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020, 21(1):180.

4. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P: Efficient Inference of Recent and Ancestral Recombination within Bacterial Populations. *Mol Biol Evol* 2017, 34(5):1167-1182.

5. Didelot X, Wilson DJ: ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 2015, 11(2):e1004041.

6. Martin-Rodriguez AJ, Higdon SM, Thorell K, Tellgren-Roth C, Sjoling A, Galperin MY, Krell T, Romling U: Comparative Genomics of Cyclic di-GMP Metabolism and Chemosensory Pathways in Shewanella algae Strains: Novel Bacterial Sensory Domains and Functional Insights into Lifestyle Regulation. *mSystems* 2022, 7(2):e0151821.

7. Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, Rooney AP, Brambilla E, Connor N, Ratcliff RM *et al*: Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci U S A* 2008, 105(7):2504-2509.

8. Wood JM, Becraft ED, Krizanc D, Cohan FM, Ward DM: Ecotype Simulation 2: An improved algorithm for efficiently demarcating microbial species from large sequence datasets. *bioRxiv* 2020:2020.2002.2010.940734.

*[Remote supervision possible]*

**28. Metabolic Digital Twins; creating personalised metabolic models using data from wearable sensors**

**Supervisor:** Shauna O’Donovan, Computational Biology Group, Department of Biomedical Engineering, Eindhoven University of Technology, [s.d.odonovan@tue.nl](mailto:s.d.odonovan@tue.nl)

The manifestation of metabolic deteriorations that accompany overweight and obesity can differ greatly between individuals, giving rise to a highly heterogeneous population. This inter-individual variation can impede both the provision and assessment of nutritional interventions as multiple aspects of metabolic health should be considered at once. Recent advances in wearable sensory technology means that is it now possible to monitor glucose excursions in free living conditions, providing unprecedented insight into an individual’s metabolic resilience. However, how to quantify meaningful features of metabolic health from these extensive time series of data still presents many challenges. Physiology-based mathematical models (PBMMs) consist of systems of ordinary differential equations that can be used to simulate the dynamics of biological systems *in silico*. Numerous PBMMs have been developed to study how the interactions between glucose, insulin, and other metabolites and hormones change during the development of insulin resistance and Type 2 Diabetes Mellitus. In previous work, we have shown that these PBMMs can be personalised by estimating a patient-specific set of model parameters from experimental data [1]. More recently we have shown that these personalised parameter estimates quantify features of metabolic resilience such as insulin resistance, liver fat, and beta-cell functionality from meal response data, providing us with greater insight into the interindividual variation in metabolic health [2]. However, these parameters are estimated using plasma concentrations of glucose, insulin, and lipids which necessitates the collection of multiple venous samples which can be invasive and cumbersome for patients. In this project, we aim to extend our parameter estimation protocol to data from wearable continuous glucose monitors. In this way, we aim to create an *in silico* simulation model that can be personalised and updated with real-time sensor data, producing a metabolic digital twin. Guidance will be provided on working with ODE models in MATLAB or Julia.

Topics: Parameter estimation, ordinary differential equations, continuous glucose monitoring, digital twins, MATLAB or Julia

1. Erdős B, van Sloun B, Adriaens ME, O'Donovan SD, Langin D, Astrup A, Blaak EE, Arts ICW, van Riel NAW. Personalized computational model quantifies heterogeneity in postprandial responses to oral glucose challenge. PLoS Comput Biol. 2021 Mar 31;17(3):e1008852. doi: 10.1371/journal.pcbi.1008852.

2. O'Donovan SD, Rundle M, Thomas EL, Bell JD, Frost G, Jacobs DM, Wanders AJ, de Vries R, Mariman ECM, van Baak MA, Sterkman L, Nieuwdorp M, Groen AK, Arts ICW, van Riel NAW, Afman LA. Quantifying the effect of nutritional interventions on metabolic resilience using personalised computational models. iScience. 2024 Feb 28; 109362. doi: 10.1016/j.isci.2024.109362.

*[Remote supervision possible]*

**29. Fuelling the fire; how does diet impact immune metabolism and chronic inflammation?**

**Supervisor:** Shauna O’Donovan, Computational Biology Group, Department of Biomedical Engineering, Eindhoven University of Technology, [s.d.odonovan@tue.nl](mailto:s.d.odonovan@tue.nl)

Cells of the innate immune system such as monocytes and macrophages form part of our first line of defence against infection. In overweight and obese states we see impairments in the functioning of the innate immune cells that play a key role in the development of atherosclerotic plaques and the chronic low-grade inflammation which is associated with the development of Type 2 Diabetes Mellitus. In this project, we aim to construct a dynamic metabolic reconstruction of a monocyte, allowing us to study the impact of diet on the metabolism and function of innate immune cells. We will use gene expression data measured in immune cells isolated from overweight and obese donors both in a fasting state and also after consumption of high carbohydrate and high-fat meals. Using constraint-based modelling context-specific metabolic reconstructions will be generated for each individual before and after the meal. These models can then be compared to identify key reactions and pathways that have changed due to the hyper-nutrient environment following a meal. These key regulatory reactions can then be incorporated into the fasting context-specific genome-scale metabolic model as logic-gates, creating ametabolic reconstruction that can simulate the dynamic response to a meal. Guidance will be provided on working with the COBRA toolbox.

Topics: Systems biology and metabolic disease, omics data integration, genome-scale metabolic models, COBRA Toolbox, MATLAB or Python.

*[Remote supervision possible]*

**30. Development and Testing of Machine Learning Framework for Effective Application on Microbiome-Based Classification Problems**

**Supervisor:** Ciara O’Donovan, Shriram Patel, SeqBiome Ltd., [ciaraodonovan@seqbiome.com](mailto:ciaraodonovan@seqbiome.com)

NGS (Next Generation Sequencing) generated datasets pose numerous analysis challenges, such as noise, high dimensionality and small sample sizes, sparsity, and intercorrelated or redundant features. This often leads to overfitting and poor generalization of the generated model. Moreover, implementing a Machine Learning (ML) technique including preprocessing of data, model selection, and performance estimation of model for microbiome data predictive analysis can be time-consuming, confusing, and difficult.

SeqBiome has developed R-based workflow that implements various ML techniques such as regression, support vector machines, decision trees, random forest and gradient-boosted trees for classification of microbiome datasets. The MSc student would be involved in code validation, testing and further development of the ML workflow using both simulated and real-world datasets. The MSc student would compare the various ML approaches on simulated and real-world datasets with focus on their performance in disease diagnosis and biomarker discovery. The MSc student would be involved in the data collection, processing, and ML implementation across projects.

*[Remote supervision possible]*

**31. Safeguarding the Chocolate Industry from the Threats Posed by Climate Change.**

**Supervisor:** James Richardson, School of Biological, Earth and Environmental Sciences; Flavia Pezzini and Catherine Kidner, Royal Botanic Garden Edinburgh, UK; Xavier Argout CIRAD, Montpelier, [JRichardson@ucc.ie](mailto:JRichardson@ucc.ie)

Climate change is threatening our ability to produce enough food. The chocolate industry, estimated to be worth US$190 billion by 2026, supports more than five million farmers globally and is facing an ever-increasing demand. *Theobroma cacao*, the source of chocolate, is a rain forest species and in general does not resist even short dry seasons. It is predicted that areas suitable for cacao cultivation may show a 73.2% decrease in range due to climate change. This project will be integral to a research program that investigates the capacity of Cacao Wild Relatives (CWRs) to improve and safeguard productivity. *Theobroma cacao* is a representative of a diverse group of species in the tribe Theobromateae that are found in diverse biomes with different adaptations to varied climate. *Theobroma* and *Herrania* have about 20 species each, which grow in Neotropical rain forest. *Guazuma* consists of two or three species that grow in both wet and dry forests of Latin America. *Glossostemon* is a genus of a single species that is found in the deserts of the Arabian Peninsula. Many CWRs in this group are thus already adapted to extreme climates. Transcriptomes of *Guazuma ulmifolia*, a dry forest species, have been generated for plants under normal or drought stressed conditions. Many of the thousands of genes that are expressed have been associated with drought tolerance or form part of the drought stress response. This project will determine whether these genes have orthologues in cacao and how different they are in sequences. This will be achieved by blasting the drought related genes in *Guazuma* against hundreds of high-quality cacao genomes that have been generated by partners in CIRAD, France and in Latin America. The aim is to understand the nature of drought tolerance in *Guazuma ulmifolia* to determine the feasibility of producing a drought tolerant cacao that will safeguard the cacao industry against the threat of climate change.

Details of the overall research program can be found at <https://ucc.ie/en/cacaowire/>

**32. Metabolic potential of early gut colonisers in early life**

**Supervisor:** Dr Susan Joyce, School of Biochemistry and Cell Biology, [S.Joyce@ucc.ie](mailto:S.Joyce@ucc.ie)

Early life gut colonising bacteria have focused on lactic acid bacteria (LAB) *Lactobacilii* and *Bifidobacteria* species1-2. However, initial colonisers are *E. coli* species followed by *Bacteroides* species, and Enterococcal species, before optimal succession by LAB occur4-8. These are not sufficiently investigated in the extremes of life (early and late). We have investigated and isolated microbial strains from 39 infants over 12 weeks from birth. Through selection we have isolated early colonizing *E. coli* species and *Bacteroides* species. We have delimited the representation of over 4,000 bacteria CFUs to just 20 clades on the basis of RAPD analysis. We then selected 2 genomes per clade for whole genome sequencing. *Therefore, 40 genomes across bacterial species, with some newly isolated species are completed and awaiting investigation.* We expect that these sequences will provide the keys to core gut colonization and reveal essential pathways in early colonization and colonization resistance of the early infant gut as well as providing unique shared traits that are important for early life from both the gut homeostasis and microbial colonization point of view.

In tandem, we have metabolically interrogated mass spectrometry profiles of all of these strains. A new programme has recently been applied to map the metabolic potential of single microbes by the Dorristein group – with whom we are now collaborating. This represents a unique opportunity for cutting edge contributions in this area based on our work and using theirs 9

This project aims are two-fold:

1. To scaffold and represent a pipeline for genome investigations of core and pan elements.
2. To assess and to predict metabolic functions using recent Nature publication microbeMASST9: a taxonomically informed mass spectrometry search tool for microbial metabolomics data

The project is expected to yield data on:

1. overlapping homologous regions relative to existing early gut colonizer frameworks
2. To identify core and differential genetic representation over the time frame for 1 week to 3 months of life.
3. To examine the homologues here that may identify early and important metabolism including bile acid species using the Heinekin paper 20194 as a template to map representation at 3 months of life
4. To uncover important metabolising gene representation in early life

References

1. Benedetta R *et al*., 2021 Early Life Microbiota Colonization at Six Months of Age: A Transitional Time Point Frontiers in Cellular and Infection Microbiology 11 https://www.frontiersin.org/article/10.3389/fcimb.2021.590202 DOI:10.3389/fcimb.2021.590202
2. Ferretti, P. *et al.* Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe* 24, 133–145 (2018).
3. Heinekin *et al.,* 2019 Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. *Microbiome* 7, 75 (2019). https://doi.org/10.1186/s40168-019-0689-3
4. Laforest-Lapointe I., et al.2017 Patterns of Early-Life Gut Microbial Colonization during Human Immune Development: An Ecological Perspective Frontiers in Immunology 8 https://www.frontiersin.org/article/10.3389/fimmu.2017.00788 DOI=10.3389/fimmu.2017.00788
5. Pattaroni, C., *et al.,* 2018 Early-Life Formation of the Microbial and Immunological Environment of the Human Airways, Cell Host & Microbe,Volume 24, Issue 6, 2018,

Pages 857-865.e4, ISSN 1931-3128,https://doi.org/10.1016/j.chom.2018.10.019.

(<https://www.sciencedirect.com/science/article/pii/S1931312818305572>)

1. Stewart, C. J. *et al,.* Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562, 583–588 (2018).
2. Vatanen, T. et al. Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. *Nat. Microbiol*. 4, 470–479 (2018).
3. Yassour, M. *et al*., Strain-level analysis of mother-to-child bacterial transmission during the first few months of life. *Cell Host Microbe* 24, 146–154 (2018).
4. Zuffa *et al.,* (2024) microbeMASST: a taxonomically informed mass spectrometry search tool for microbial metabolomics data. Nature Microbiology | Volume 9 | February 2024 | 336–345 <https://www.nature.com/articles/s41564-023-01575-9>.