Genome Data Science and AI for Therapeutic Target Discovery

Project Catalogue 2026

Project 1

Harnessing AI and multi-omics in the UK Glaucoma BioResource for therapeutic target discovery in glaucoma scarring and blindness

Supervisor 1: Dr Cynthia Yu-Wai-Man, Department of Ophthalmology, School of Life Course & Population Sciences, Faculty of Life Sciences & Medicine
Supervisor 2: Dr Mario Falchi, Twin Research and Genetic Epidemiology, School of Life Course and Population Sciences, Faculty of Life Sciences & Medicine

Contact: cynthia.yu-wai-man@kcl.ac.uk and mario.falchi@kcl.ac.uk

Project Abstract: Glaucoma blinds millions of people worldwide and scarring is the main reason why surgery in glaucoma stops working. There is a large unmet clinical need to develop a safe and targeted therapeutic to prevent scarring and blindness in glaucoma. Over the past decades, clinical trials have used therapeutic targets identified from animal studies and small human studies, but these clinical trials have failed to show efficacy in glaucoma patients. Little is currently known about the mechanisms and therapeutic targets in patients in glaucoma scarring and blindness.

The UK Glaucoma BioResource is a disease-specific, large-scale biobank of over 3000 blood and tissue samples and detailed phenotypic data from glaucoma surgery patients, collected from 7 tertiary referral hospitals and 25 glaucoma specialists in the UK. This project leverages the UK Glaucoma BioResource to identify genetic and surgical factors affecting the surgical outcome and blindness in glaucoma patients. Using machine learning models, including deep learning approaches, surgical failure and blindness will be classified from a large, combined dataset of 28,000 RNA-Sequencing and multiplex cytokines analyses, 2,000 longitudinal surgical bleb eye photos, and deep phenotyping from glaucoma patients.

The key aims and objectives are to: (1) adapt an AI model to classify the risk of glaucoma scarring and blindness, (2) identify the genetic determinants and mechanisms, (3) assess the relative impact of genetics and surgical factors on glaucoma scarring and blindness, and (4) integrate multi-omics data to discover predictive signatures, therapeutic targets and modifiable risk factors.

Project 2

AI-driven integrative multiomics analysis of chromatin-regulator neurodevelopmental disorders for therapeutic target discovery

Supervisor 1: Dr Cristina Dias, Department of Medical and Molecular Genetics, School of Basic & Medical Biosciences, Faculty of Life Sciences & Medicine
Supervisor 2: Dr Sophia Tsoka, Department of Informatics, Faculty of Natural, Mathematical and Engineering Sciences

Contact: Cristina.dias@kcl.ac.uk and sophia.tsoka@kcl.ac.uk

Project Abstract: Chromatin-regulator neurodevelopmental disorders (NDDs), or chromatinopathies, are a rapidly expanding group of conditions caused by mutations in genes that control how DNA is packaged and expressed. More than 100 chromatinopathies are now recognised, affecting an estimated 1 in 200–300 individuals. They are among the most common genetic causes of intellectual disability and autism and often involve additional medical complications.

Although these disorders are caused by different genes, they show striking similarities in clinical features and molecular mechanisms, suggesting that they disrupt shared biological pathways. Understanding how these distinct mechanisms give rise to overlapping symptoms could uncover new therapeutic opportunities that benefit multiple disorders.

This PhD project will apply artificial intelligence (AI) and advanced computational modelling to integrate multi-omic data from patient cohorts and model system publicly available datasets. By analysing these complex datasets together, the student will:

1. Identify convergent molecular pathways that are commonly affected across chromatinopathies;
2. Stratify disorders based on molecular mechanisms and chromatin state; and
3. Prioritise therapeutic targets and candidate compounds using AI-driven drug–target mapping and repurposing tools.

The project combines expertise in chromatin biology and neurodevelopment (Dias Lab) with cutting-edge AI and systems biology (Tsoka Lab). It will deliver the first AI-based integrative framework for chromatinopathies, revealing druggable molecular networks and establishing a foundation for precision therapeutic strategies in neurodevelopmental disease.
Because similar chromatin-regulating genes are also implicated in cancer, the findings will provide insights relevant to both developmental and oncogenic contexts. The computational approaches developed will be broadly applicable to other rare and epigenetic disorders, contributing to the advancement of AI-enabled therapeutic discovery in genomics.

Project 3

Lab-in-the-Loop AI for Transcription Factor Variants in Human Pancreas Development

Supervisor 1: Dr Xianghua Li, Medical and Molecular Genetics, BMBS, FolSM
Supervisor 2: Dr Rocio Sancho, Centre for Gene Therapy & Regenerative Medicine

Contact: xianghua.li@kcl.ac.uk and rocio.sancho@kings.ac.uk

Project Abstract: Transcription factors (TFs) are key regulators of gene expression, controlling cellular identity and function. Disruptions in TF activity drive a range of diseases, including diabetes and cancer. However, the computational challenge of predicting how specific mutations impact TF function remains a major hurdle, particularly for proteins with large intrinsically disordered regions. A significant proportion of TF variants are classified as variants of uncertain significance (VUS), with many TFs deemed “undruggable,” limiting therapeutic opportunities.

This PhD project will integrate AI-driven modelling with experimental validation in a lab-in-the-loop framework to predict, measure, and therapeutically target the effects of TF mutations. At the heart of this project is the TF-MAPS platform, developed by Dr Xianghua Li, which uses deep mutational scanning (DMS) to systematically assess how mutations impact TF abundance, stability, and DNA binding. The student will apply this high-throughput platform to the NGN3 transcription factor, a key player in pancreatic development.

By coupling DMS data with multi-omics datasets and applying AI models (e.g., deep learning, generative modelling), the student will map the functional consequences of TF mutations and predict how they disrupt transcriptional networks. These computational insights will inform targeted experimental validation using iPSC-derived pancreatic organoids, led by Dr Rocio Sancho, to assess how NGN3 and other TF mutations affect pancreatic differentiation, β-cell function, and disease phenotypes.

In an iterative lab-in-the-loop approach, the student will refine AI models based on experimental data, using feedback from the wet-lab results to improve predictions and generate genetic interventions to restore TF activity. This project exemplifies the integration of AI and experimental biology, aligning with the programme's goal of modernising therapeutic target discovery.

Project 4

Decoding Shared Pathways of Messenger RNP Dysfunction in Neurodegeneration

Supervisor 1: Prof Jernej Ule, Basic and Clinical Neuroscience (BCN), School of Neuroscience, IoPPN
Supervisor 2: Dr Sarah Marzi, Basic and Clinical Neuroscience (BCN), School of Neuroscience, IoPPN

Contact: jernej.ule@kcl.ac.uk and sarah.marzi@kcl.ac.uk

Project Abstract: Messenger ribonucleoprotein complexes (mRNPs), comprising mRNA and associated proteins, serve as the core units governing mRNA regulation and cellular function. Dysregulation of these complexes is a hallmark of various human diseases, including cancer and neurodegenerative disorders such as Amyotrophic Lateral Sclerosis (ALS), often stemming from mutations in RNA-binding proteins or repeat expansions that impair RNA processes. Historically, insights into how human genetic variation influences mRNP assembly and function have been constrained, primarily relying on data from a handful of cancer cell lines, limiting applicability to disease-relevant cell types. Recent advancements now enable the development of sophisticated machine learning models to predict the effects of genetic variants on mRNP formation and pathological transcriptomic outcomes. These include the aggregation and standardization of extensive CLIP-Seq datasets, innovative techniques for large-scale profiling of protein-RNA and RNA-RNA interactions, and adaptable sequence-to-function model architectures that can be refined with experimental data.

This study aims to construct multimodal sequence-to-function models that simulate mRNP assembly and key regulatory processes, such as splicing, polyadenylation, gene expression, and ribosome occupancy. These models will be applied to investigate the consequences of rare genetic variants identified in large-scale datasets from initiatives focused on neurodegenerative diseases. By integrating genomic, transcriptomic, and rich clinical data from these resources, the research will examine correlations between mRNP-disrupting mutations, disease progression, and altered gene expression profiles. To validate model predictions, selected mutations will be experimentally assessed using isogenic induced pluripotent stem cell (iPSC) lines, differentiated into neurons to mimic disease-relevant contexts. This approach promises to uncover convergent mechanisms of mRNP dysregulation in neurodegeneration, potentially identifying novel therapeutic targets and enhancing personalized medicine strategies for affected individuals.

Project 5

Uncovering the role of transfer RNA in human cancers

Supervisor 1: Dr Michelle Holland, Medical and Molecular Genetics, School of Basic and Medical Biosciences, FoLSM
Supervisor 2: Dr Richard Dillon, Medical and Molecular Genetics, School of Basic and Medical Biosciences, FoLSM

Contact: michelle.holland@kcl.ac.uk and richard.dillon@kcl.ac.uk

Project Abstract: Transfer RNAs (tRNAs) perform a critical function in gene expression as adapter molecules required for the translation of the nucleic acid sequence of messenger RNA (mRNA) into the amino acid sequence of the cognate protein. There are more than 600 predicted tRNA genes in humans which have only recently been found to be dynamically regulated in tissue and disease contexts. The multicopy nature of tRNAs has historically posed technical challenges for investigating the extent of genetic variation of tRNA genes at population scale as it requires whole genome sequencing (WGS) data to be available for large cohorts. However, small-scale studies have suggested that both copy number variation and sequence variation are common. Such genetic variation could influence tRNA functions in mRNA translation and beyond through influencing abundance and function by altering the tRNA modification profile or otherwise. Changes in tRNA abundance have recently been recognised as a common feature of many cancers, suggesting that they could be a potential therapeutic target. The aim of this project is to utilise the WGS data available from The Cancer Programme of the 100,000 Genomes Project to gain an understanding of the breadth of germline variation at tRNA genes in humans and to identify any common signatures of how this may be altered in the context of cancer. To understand how genetic variation links to tRNA abundance and modification profile, direct RNA sequencing will be employed in relevant models. Finally, functional impact will be defined through targeted genetic approaches and small molecule disruption. This work sets out to identify if tRNAs are a tractable therapeutic target in cancers.

Project 6

Human Genetic Evaluation of Mitochondrial Transcription as a Metabolic Drug Target

Supervisor 1: Dr Alan Hodgkinson, Medical and Molecular Genetics, Basic and Medical Biosciences, FoLSM
Supervisor 2: Prof Afshan Malik, Department of Diabetes and Obesity, School of Cardiovascular and Metabolic Medicine & Sciences, FoLSM

Contact: alan.hodgkinson@kcl.ac.uk and afshan.malik@kcl.ac.uk

Project Abstract: Mitochondria are essential for energy production and metabolic regulation in human cells. Their activity depends on coordinated expression of genes encoded in both the nuclear and mitochondrial genomes. In the liver, changes in mitochondrial DNA (mtDNA) transcription directly affect oxidative phosphorylation (OXPHOS), influencing how cells use fuels such as fats and carbohydrates. Recent studies in mice have shown that pharmacological inhibition of mtDNA transcription in the liver can improve metabolic health, reducing liver fat and restoring glucose tolerance. However, it is unknown whether similar effects occur in humans.

This PhD project will use human genetic and transcriptomic data to model the effects of reduced mtDNA transcription in the liver, providing a “natural experiment” to explore how mitochondrial gene expression influences metabolism and disease risk. The student will integrate large-scale RNA sequencing and genome-wide genetic data to (1) quantify mtDNA transcription in human liver tissue and identify its nuclear genetic regulators, (2) develop predictive models that estimate mtDNA expression levels from genetic data using machine learning, and (3) apply these models to population-scale cohorts such as UK Biobank to test causal effects on metabolic traits, liver health and disease outcomes using Mendelian Randomisation and related approaches.

This work will generate the first comprehensive genetic framework for understanding mitochondrial transcriptional regulation in human liver and its impact on systemic metabolism. By revealing how naturally occurring genetic variation in mitochondrial function shapes risk for metabolic diseases such as fatty liver and type 2 diabetes, this project will provide new insights into the potential of targeting mitochondrial transcription as a therapeutic strategy. The research combines computational genomics, statistical genetics and human physiology, offering training at the interface of genetics, bioinformatics and translational biomedical science.

Project 7

Genetic determinants of multi-fluid metabolomics and links to cardiometabolic, hepatic, and renal traits

Supervisor 1: Dr Cristina Menni, Department of Twin Research & Genetic Epidemiology, FoLSM
Supervisor 2: Prof Kerrin Small, Department of Twin Research & Genetic Epidemiology, FoLSM

Contact: cristina.menni@kcl.ac.uk and kerrin.small@kcl.ac.uk

Project Abstract: Metabolomics provides a direct snapshot of biochemical activity, reflecting disease mechanisms, lifestyle, environmental influences, microbiome function, and pharmacological effects. Its integration with human genetics offers powerful opportunities to uncover pathways linking genotype to phenotype and to identify biomarkers relevant to disease risk and treatment response.

Genome-wide studies have mapped genetic variants influencing metabolites in plasma, urine, saliva, and stool, highlighting the genetic contribution to metabolic traits. However, each biofluid provides unique biological information, and limiting studies to serum or plasma overlooks key compartment-specific processes.

We have generated extensive metabolomics data using the largest Metabolon panel across serum, urine, saliva, and stool, together with detailed clinical phenotypes from nearly 4,000 TwinsUK participants. This unique multi-fluid dataset enables systematic exploration of genetic regulation and clinical associations of metabolites across biofluids.

This PhD project will (i) identify genetic determinants of metabolites within each biofluid through large-scale genome-wide association studies and (ii) examine relationships between metabolites and clinical traits relevant to cardiometabolic, hepatic, and renal health. Integrating these results will reveal shared and distinct metabolic pathways across biological compartments and identify candidates for therapeutic targeting.

This project will create a comprehensive, open-access resource integrating genetic, metabolic, and clinical data to advance data-driven identification of therapeutically relevant targets. The studentship will provide training in statistical genetics, bioinformatics, AI, and systems biology, equipping the student to drive innovation in genetics-based drug discovery and precision medicine.

Project 8

Genetics-Guided Discovery and Functional Validation of Hippocampal Therapeutic Targets for Bipolar Disorder

Supervisor 1: DrTimothy Powell, Social, Genetic & Developmental Psychiatry Department, Mental Health & Psychological Sciences, IoPPN
Supervisor 2: Prof Professor Sandrine Thuret, Basic and Clinical Neuroscience, School of Neuroscience, IoPPN

Contact: timothy.1.powell@kcl.ac.uk and sandrine.1.thuret@kcl.ac.uk

Project Abstract: Bipolar disorder (BD) is a severe psychiatric condition characterised by mood instability, cognitive difficulties, and structural changes in the brain. One of the most consistent biological findings in BD is reduced hippocampal volume, which is closely linked to cognitive impairment and illness burden. Notably, lithium—the most effective treatment for BD—is associated with increased hippocampal volume in patients and enhanced function of hippocampal progenitor cells, making this brain region an important focus for therapeutic discovery.

This project aims to identify hippocampal-expressed genes that contribute to genetic risk for BD and to determine whether their cellular effects can be reversed by lithium or other repurposed compounds. The student will integrate BD genome-wide association study (GWAS) results with hippocampal transcriptomic datasets using methods including transcriptome-wide association studies (TWAS), colocalisation, and statistical fine-mapping. To understand how risk-associated genes operate within broader biological systems, the student will apply network-based approaches (e.g., WGCNA) to full hippocampal transcriptomes, allowing identification of enriched co-expression modules and genetically regulated hub genes.

To explore therapeutic opportunities, the student will use computational drug repurposing resources such as the Connectivity Map platform, which matches disease-associated gene expression signatures to compounds predicted to reverse them. This will generate a shortlist of repurposing candidates for experimental testing.

The strongest targets will then be functionally validated in human hippocampal progenitor cells using CRISPR interference or activation. The student will examine effects on proliferation, differentiation, and cell death across neuronal and glial lineages—key processes that shape hippocampal structure and function. Lithium and computationally predicted compounds will be tested for their ability to rescue adverse cellular phenotypes.

This interdisciplinary project combines human genetics, multi-omics integration, computational drug repurposing, and functional genomics to uncover novel therapeutic pathways in bipolar disorder.

Project 9

Multimodal AI for predicting personalised, genotype-specific therapeutic response in cancer and rare disorders

Supervisor 1: Dr Christopher Banerji, Comprehensive Cancer Centre, School of Cancer and Pharmaceutical Sciences, FoLSM
Supervisor 2: Prof Peter Zammit, Randall Centre for Cell & Molecular Biophysics School of Basic & Medical Biosciences, FoLSM
Supervisor 3: Prof Anita Grigoriadis, Comprehensive Cancer Centre, School of Cancer and Pharmaceutical Sciences, FoLSM

Contact: christopher.banerji@kcl.ac.uk, peter.zammit@kcl.ac.uk, anita.grigoriadis@kcl.ac.uk

Project Abstract: Variation in drug response remains a major challenge for precision medicine. Therapies that are highly effective in some patients can be ineffective or even harmful in others with the same diagnosis. Genetic differences often drive this heterogeneity, yet such effects are usually discovered only after clinical deployment, resulting in avoidable morbidity and failed clinical trials. There is an urgent need for predictive frameworks that can anticipate treatment response before therapy begins.

This project will develop an AI-based system that models how both genetic variants and drugs reshape cellular gene-expression programs, providing a unified approach to genotype-specific therapy prediction for cancer and rare diseases.

The student will undertake 3 linked work-packages (WPs):

WP 1 will create two nested predictive models: one that learns how genetic perturbations shift transcriptomic states using large datasets such as LINCS L1000 and sci-Plex, and another that predicts how drug structure and dose alter those states, using deep-learning methods and chemical embeddings (SMILES). Combining these models enables in silico matching of drugs that reverse mutation-induced expression changes and discovery of new therapeutic candidates.

WP 2 links these predicted transcriptomic shifts to outcome in cancer and rare diseases, testing whether they move tumours or tissues toward improved survival or reduced disease activity. Measurable read-outs of therapeutic benefit for cancer will include established biomarkers such as signalling entropy. The rare disease that we have selected is the muscle wasting disease facioscapulohumeral muscular dystrophy (FSHD) for which we have devised and verified molecular biomarkers of disease progression.

WP 3 extends this framework to the clinic by integrating a generative histopathology model that infers transcriptomic profiles directly from routine H&E images, enabling high-throughput patient-level therapy simulation without new sequencing data.

Together, these components form a scalable, interpretable platform linking genotype, transcriptome and phenotype to guide drug discovery and personalised treatment design.

Project 10

Genetic bases of lifetime disease burden

Supervisor 1: Dr Maxim Freydin, Twin Research and Genetic Epidemiology, Life Course and Population Sciences, FoLSM
Supervisor 2: Prof Claire Steves, Twin Research and Genetic Epidemiology, Life Course and Population Sciences, FoLSM

Contact: maxim.freydin@kcl.ac.uk and claire.j.steves@kcl.ac.uk

Project Abstract: Natural history of human disease involves periods of health, acute disease, chronic disease, and multimorbidity states. Individual trajectories of this health/disease evolution depend on genetic, social, and environmental factors, as well as poorly understood relationships between diseases manifested as comorbidity and inverse comorbidity. One of the outcomes of these relationships is the total number of diseases a person would experience during the life course (lifetime disease burden, LDB). Understanding genetic underpinnings of LDB might be a step forward in developing individualized prediction and prevention strategy. Yet, this phenomenon has never been explicitly studied.

The study aims at exploring molecular factors underlying LDB using genetic and multi-omics approach. Using UK Biobank as the primary source, we will carry out genome-wide association studies (GWAS) for LDB measure as the sum of ICD10 disease classification codes followed by an intensive post-GWAS analyses including gene mapping and prioritization, genetic correlation analysis, pathway analyses, colocalization with eQTLs and mQTLs, and tissue specific heritability analysis. To establish possible causal links between LDB and potential risk factors, we will be applying Mendelian Randomization. Polygenic risk scores will be developed to facilitate prediction of LDB using additional datasets, such as TwinsUK. Further mechanistic insight into LDB will be obtained via metabolomic and proteomic analyses.

This study introduces the concept of LDB as a quantitative trait analyzable via GWAS, something that has not previously been explicitly defined or investigated, which allows more holistic assessment of an individual's lifelong health trajectory rather than focusing on single diseases or comorbidity clusters. Its importance is defined by the potential to uncover the genetic and molecular architecture of LDB which can reveal shared biological mechanisms underlying multimorbidity, enable earlier identification of individuals at high lifetime risk, and ultimately support the development of personalized healthcare strategies.

Project 11

A Genome-Scale Foundational Model Integrating DNA Language Models and GNNs for Causal Inference in Human Genetics

Supervisor 1: Prof Michael Simpson, Department of Medical and Molecular Genetics, Faculty of Life Sciences & Medicine
Supervisor 2: Dr Aleksej Zelezniak, Randall Centre for Cell & Molecular Biophysics, Faculty of Life Sciences & Medicine

Contact: michael.simpson@kcl.ac.uk and aleksej.zelezniak@kcl.ac.uk

Project Abstract: Understanding how genetic or molecular perturbations propagate through biological systems remains a central challenge in therapeutic target discovery. While large-scale human genetics offers thousands of naturally occurring perturbations linked to disease outcomes, current machine-learning approaches struggle to turn these associations into mechanistic insight. Most existing models identify statistical patterns but operate as opaque predictors, offering a limited understanding of how a variant or drug influences cellular pathways. This limits our ability to discover causal drivers of disease and to identify precise intervention points.

This project will develop a new computational framework that moves beyond predictive modelling towards mechanistic and generative reasoning. The aim is to engineer an AI system capable of explaining how perturbations produce downstream effects by learning the underlying structure of biological networks. The first stage will focus on building a large-scale, heterogeneous biological knowledge graph that integrates genetic, transcriptomic, proteomic and regulatory information. This graph will serve as the foundation for a graph neural network (GNN) trained to represent the core interactions governing cellular function.

The novelty of the work lies in a two-stage "in silico perturbation engine". After training the GNN to capture the system's baseline dynamics, a generative agent inspired by Generative Flow Networks will be used to explore the model's learned representations. This agent will generate diverse, step-by-step causal paths that explain how a specific perturbation leads to a given outcome, producing testable mechanistic hypotheses rather than opaque predictions.

By combining genome-scale genetic signals with advanced GNN and generative modelling, this project will create a platform to map causal consequences of perturbations and highlight optimal targets for intervention. The resulting framework has the potential to improve the mechanistic interpretability of genetic associations and accelerate the discovery of therapeutically actionable pathways.

Project 12

Decoding the Heart–Brain Axis for Therapeutic Discovery Using Multi-Omics, Machine Learning, and Causal Genetics

Supervisor 1: Dr Pier-Giorgio Masci, Cardiovascular Imaging, Biomedical Engineering and Imaging Sciences, Faculty of Life Sciences and Medicine
Supervisor 2: Dr Maxim Freydin, Twin Research and Genetic Epidemiology, Life Course and Population Sciences, Faculty of Life Sciences and Medicine

Contact: pier_giorgio.masci@kcl.ac.uk and maxim.freydin@kcl.ac.uk

Project Abstract: Background. Cardiovascular disease and dementia are two of the most pressing global health challenges. They frequently co-occur in the same patients and share metabolic, vascular, and inflammatory pathways, indicating a biological "heart–brain axis". However, the molecular mechanisms underlying this axis remain poorly characterised. Leveraging machine learning applied to large-scale genomics and proteomics offers an unprecedented opportunity to map these mechanisms and identify druggable targets.

Aim: To decode the molecular architecture of the heart–brain axis through integrated multi-omics and machine-learning approaches, with the overarching scope to identify novel therapeutic pathways for tackling cognitive decline and dementia.

Objectives:

1. Molecular endotype discovery: using multi-omics data from UK Biobank and TwinsUK, we will identify cardiovascular or vascular molecular subtypes ("endotypes") that associate with accelerated brain ageing.

2. Generative modelling of the heart–brain axis: we will build a computational framework to jointly capture molecular profiles in heart and brain, revealing shared ageing signatures (endophenotypes) across these systems.

3. Causal target prioritisation: through genome-wide association studies, fine-mapping, and Mendelian randomisation, we will identify candidate credible genes and proteins that causally mediate heart–brain interactions, and then prioritise targets for intervention by cross-referencing with drug–gene databases and functional perturbation datasets.

Novelty and Importance: This project will systematically combine multi-omics profiling, causal genomics, and advanced machine-learning to map druggable pathways across the cardiovascular–neurodegenerative ageing axis. By moving beyond traditional phenotypes and focusing on molecular mechanisms, it aims to generate actionable insights into novel therapeutic targets and repurposing opportunities.

Project 13

A multimodal and interaction-based approach to the multi-omics study of the biological basis of neurodegenerative diseases

Supervisor 1: Dr Alfredo Iacoangeli, Biostatistics and Health Informatics, MHaPS, IoPPN
Supervisor 2: Prof Ammar Al-Chalabi, Basic and Clinical Neuroscience, School of Neuroscience, IoPPN

Contact: alfredo.iacoangeli@kcl.ac.uk and ammar.al-chalabi@kcl.ac.uk

Project Abstract: Neurodegenerative diseases such as ALS, Alzheimer's disease, and Parkinson's disease arise from complex interactions between genetic, molecular, and environmental factors. Despite the availability of large-scale genomic and multi-omics datasets, most analytical approaches remain limited to univariate testing, assessing each biological factor independently. This strategy lacks the power to detect rare variants, weak-effect signals, or the higher-order epistatic and non-linear interactions that may jointly drive disease risk. Testing all possible interactions across millions of genomic, transcriptomic, proteomic, or methylation features is computationally infeasible and statistically penalised by extreme multiple-testing burdens. As a result, a substantial proportion of disease-relevant biology, including multi-layer interactions and mechanistic patient subgroups, remains undiscovered.

This project proposes a multimodal and interaction-based analytical framework to address these limitations and advance the understanding of disease mechanisms in neurodegeneration. By integrating genomics, transcriptomics, proteomics and methylation data with state-of-the-art machine learning, we aim to identify optimal combinations of features across data modalities that collectively explain disease risk. The project will systematically evaluate existing interaction-based methods and develop a next-generation computational approach capable of modelling non-linear, multi-omic interactions at scale. Applying this methodology to large public neurodegenerative disease datasets, we will identify biologically homogeneous patient subgroups linked to distinct mechanistic drivers and clinical outcomes. Such stratification has the potential to inform personalised diagnostic and prognostic profiles and guide targeted therapeutic development.

A further objective is to perform cross-disease analyses to uncover shared molecular mechanisms across neurodegenerative conditions, providing insight into convergent biological pathways. The project's computational pipeline will be generalisable, enabling future application to other complex diseases and biological systems.

By combining machine learning, interaction-based modelling, and multi-omics integration, this work aims to overcome current analytical limitations and deliver a scalable, data-driven framework for mechanistic discovery and patient stratification in neurodegenerative disease.

Project 14

Sex-Specific Multi-Omic Stratification of Alzheimer's Disease for AI-Driven Therapeutic Target Discovery.

Supervisor 1: Dr Ahmad Al Khleifat, Basic and Clinical Neuroscience, IoPPN
Supervisor 2: Prof Alfredo Iacoangeli, Biostatistics and health informatics, IoPPN

Contact: ahmad.al_khleifat@kcl.ac.uk and alfredo.iacoangeli@kcl.ac.uk

Project Abstract: Alzheimer's disease (AD) shows clear differences between women and men in incidence, progression, biomarker trajectories, and underlying biology. Yet most large-scale analyses treat the disease as sex-neutral, potentially masking mechanisms that are unique to, or more pronounced in, one sex. The UK Biobank (UKB) provides an exceptional opportunity to address this gap, offering whole-genome sequencing, plasma proteomics (Olink and SomaLogic), DNA methylation data, RNA, detailed lifestyle and cardiometabolic profiles, and repeated cognitive assessments for over 500,000 participants, allowing sex-specific disease mechanisms to be investigated at population scale.

This project aims to integrate these diverse data layers to identify sex-specific, biologically meaningful subtypes of Alzheimer's disease risk and progression. Using advanced unsupervised machine-learning approaches, the project will separately and jointly model molecular and phenotypic variation in women and men, combining genetic variation, proteomic signatures, methylation, imaging-derived features, and longitudinal cognition. The goal is to uncover robust sex-aware molecular patterns that explain why AD manifests differently across sexes.

Each subtype will be characterised through pathway and network analyses, as well as cell-type deconvolution, to identify sex-dependent biological processes, such as microglial activation states, lipid metabolism and APOE-related mechanisms, synaptic resilience, vascular involvement, and tau-associated pathways. Subtypes will then be evaluated for their association with sex-specific trajectories of cognitive decline, imaging progression, biomarker changes, and incident dementia diagnoses captured through linked health records.

A key innovation is the development of sex-specific drug-repurposing predictions. Molecular signatures will be matched to drug-perturbation datasets (e.g., LINCS L1000) and mechanistic drug–target resources to identify compounds most likely to modulate pathways relevant to each sex-specific subgroup. Overall, this PhD will deliver the first population-scale, sex-aware, multi-omic stratification of Alzheimer's disease, offering new insights into its molecular heterogeneity and creating a foundation for precision, sex-tailored therapeutic strategies.

Project 15

Identifying molecular drivers of cardiometabolic disease via adipose tissue multi-omics integration

Supervisor 1: Prof Kerrin Small, Department of Twin Research & Genetic Epidemiology, School of Life Course and Population Sciences, FoLSM
Supervisor 2: Dr Pier-Giorgio Masci, Cardiovascular Imaging, Biomedical Engineering and Imaging Sciences, Faculty of Life Sciences and Medicine

Contact: kerrin.small@kcl.ac.uk and pier_giorgio.masci@kcl.ac.uk

Project Abstract: Cardiometabolic diseases, including Type 2 Diabetes and cardiovascular disease, are an increasing global problem. Risk of cardiometabolic disease is associated with multiple, diverse genetic factors. A key tissue mediating genetic risk of cardiometabolic disease is adipose tissue (fat). Adipose tissue is an endocrine organ which plays a central role in the regulation of metabolism, inflammation, synthesis and secretion of hormones, and serves as a buffering system for lipid energy balance. Adipose tissue function and dysfunction can differ across the body, and differential deposition of fat in different locations in the body strongly influences risk of cardiometabolic disease.

This project will focus on identifying and investigating the genetic variants that regulate processes in adipose tissue and determine fat distribution across the body in order to discover the molecular pathways and networks that underly development of cardiometabolic disease. The project will integrate genetic and GWAS data from large biobanks with a unique adipose tissue multi-omic dataset from ~800 twins, including transcriptomics, proteomics, epigenetics and fatty acid profiles, alongside matched high-dimension MRI imaging and clinical phenotypes. By integrating genetic, multi-omic, and clinical data in this key cardiometabolic tissue, the project will discover regulatory networks and molecular mechanisms that mediate the genetic risk of cardiometabolic disease, thereby identifying potential therapeutic targets.

The studentship will offer comprehensive training in complex disease statistical genetics, adipose biology, and multi-modal data analyses, preparing the student to contribute to advancements in drug discovery and precision medicine.

Project 16

Integrative Genomic Analyses to Dissect Genetic and Environmental Modifiers of APOE-linked Risk in Alzheimer's Disease

Supervisor 1: Dr Sarah Marzi, Basic and Clinical Neuroscience, IoPPN
Supervisor 2: Dr Oliver Pain, Basic and Clinical Neuroscience, IoPPN

Contact: sarah.marzi@kcl.ac.uk and oliver.pain@kcl.ac.uk

Project Abstract: Alzheimer's disease (AD) is the leading cause of dementia, and APOE4 is the strongest common genetic risk factor. Its penetrance varies across populations, highlighting the influence of genetic and environmental modifiers. Evidence suggests that vitamin D receptor (VDR) signalling may modulate APOE-associated microglial dysfunction, supported by microglial and neurovascular epigenomic analyses that predict VDR agonists as compounds capable of reversing AD risk regulatory signatures.

This PhD will identify and characterise genetic and environmental modifiers of APOE risk, combining large-scale discovery with mechanistic validation. Population-scale genetic datasets combined with environmental exposures and single-cell transcriptomic and epigenomic data from microglia, astrocytes and neurovascular cells will be analysed using genome-wide interaction scans, cross-ancestry comparisons, polygenic scores, and causal inference to discover both known and novel modifiers.

Functional validation will be performed in iPSC-derived microglia and astrocytes. CRISPR perturbations, VDR agonist treatments, multi-omic profiling, and functional assays of phagocytosis, cytokine secretion, and chemotaxis will test the impact of APOE isoforms, VDR signalling, and candidate modifiers. iPSCs from donors of diverse ancestries in village culture systems will allow modelling of gene × environment interactions, including vitamin D exposure.
The novelty of this project lies in integrating population and single-cell discovery, cross-ancestry analyses, and predictive computational modelling with mechanistic iPSC-based validation. Its importance is in uncovering both VDR-dependent and novel modulators of APOE-linked risk, generating mechanistically supported therapeutic targets and informing precision interventions for AD across diverse populations.

Project 17

AI-based Drug Discovery For Diabetic Microangiopathy

Supervisor 1: Dr Konstantinos Theofilatos, Cardiovascular, SCMMS, FoLSM
Supervisor 2: Dr Anna Zampetaki, Cardiovascular, SCMMS, FoLSM

Contact: konstantinos.theofilatos@kcl.ac.uk and anna.zampetaki@kcl.ac.uk

Project Abstract: Microangiopathy is a major complication in diabetes mellitus with a pronounced impact on microvascular perfusion, vessel rarefaction and tissue ischemia. There are very limited therapeutic options to prevent the progression of the disease and thus there exists a great need to accurately determine patients' individualized risk of progression to more severe microangiopathy and design novel therapeutic interventions. As part of our ongoing research, we have used human blood vessel organoids (BVOs), an attractive model of microvasculature, to identify novel drug targets of diabetic microangiopathy using multi-omics data (spatial transcriptomics and mass-spectrometry proteomics). This study aims to harness the power of Artificial Intelligence (AI) and machine learning to develop prognostic models for diabetic microangiopathy, stratify them on different risk groups and develop novel therapeutic methods to target them. The development of novel prognostic models will be based on whole-genome sequencing data and proteomics data from the UK Biobank population study, focusing on the 120 novel gene targets identified from our multi-omics BVO study. Machine-learning dimensionality reduction and multivariate survival models will be used to develop mortality and cardiovascular risk models. In addition, the multi-omics data of BVOs from the supervisor's lab and Human participants from the UK Biobank will be analyzed with network and proteogenomics techniques to prioritize drug targets. Finally, in silico drug screening and ligand design techniques based on artificial intelligence-based workflows will be deployed to identify novel and known compounds to target the microangiopathy targets. This project will set the basis for developing novel prognostic scores and therapeutic methods for diabetic cardiomyopathy, a condition that is currently not treated effectively. Moreover, a series of AI, machine learning and network-based workflows will be assessed and fine-tuned, significantly contributing to the improvement of in silico methods for drug targeting.

Project 18

Identification of causal factors underlying sarcopenia using AI and mathematical models for predicting molecular drivers of muscle homeostasis and repair

Supervisor 1: Dr Robert Knight, Centre for Craniofacial and Regenerative Biology, FoDOCS
Supervisor 2: Dr Saeed Shoaie, Centre for Host Microbial Interactions, FoDOCS

Contact: robert.knight@kcl.ac.uk and saeed.shoaie@kcl.ac.uk

Project Abstract: A defining feature of ageing and many diseases is impaired muscle function, resulting in weakness. This is often associated with sarcopenia, involving altered muscle metabolism, reduced myofibre capacity and a reduction in myofibre size. Identifying causal factors underlying sarcopenia that could be targeted for interventions offers a potential avenue for improving muscle strength and function across a wide range of conditions. A challenge for identifying such factors is that although we know many of the molecules that regulate muscle function, we do not know which can induce sarcopenia. In this proposal we aim to leverage human genome variant data, gene expression data and computational models of gene regulation to predict candidate causal genes and molecules for sarcopenia. This will involve comparing gene expression data to genomic variants using data from the UK Twins study and the GTEx Project to identify genes associated with sarcopenia. By mapping candidate genes associated with variants and/ or altered expression to genome-scale models of gene regulatory networks (GRNs) and metabolic models (GEMs) in a deep learning AI model, we will make predictions of molecular outcomes. Those molecules predicted to have a potentially causal role in sarcopenia will then be tested in human iPSC-derived muscle, and in vivo in zebrafish, to determine their function in maintaining myofibre function, and in muscle homeostasis and repair.

Project 19

Genetics-Informed AI Retrobiosynthesis for Therapeutic Target Discovery in Low-Grade Inflammation

Supervisor 1: Dr Miao Guo, Department of Engineering, Faculty of Natural Mathematical and Engineering Sciences
Supervisor 2: Prof Adil Mardinoglu, Faculty of Dentistry, Oral & Craniofacial Sciences, Centre for Host-Microbiome Interactions (CHMI)

Contact: miao.guo@kcl.ac.uk and adil.mardinoglu@kcl.ac.uk

Project Abstract: Low-grade systemic chronic inflammation (LGSCI) is a key contributor to metabolic dysfunction and non-communicable diseases (NCDs), yet its molecular origins remain incompletely understood. LGSCI arises from interactions between host genetic variation, metabolic status, environmental exposures, and gut microbiome–derived metabolites. Although genome-wide association studies (GWAS) have identified genetic determinants of inflammatory biomarkers, the pathways linking these variants to microbiome-modulated metabolic signals and chronic inflammation remain poorly resolved. Decoding these mechanisms is essential for identifying therapeutically actionable targets.

This PhD project will integrate human genetic data with multi-omics layers—including metabolomics, metagenomics, transcriptomics, and proteomics—to uncover genetically supported regulators of LGSCI. Using UK Biobank and other large-scale cohorts, the student will apply statistical genetics, Mendelian randomisation, colocalisation, and polygenic analyses to identify causal host genes, microbial pathways, and metabolites influencing inflammatory biomarkers and NCD risk.

Building on our existing retrobiosynthesis and cross-omics interaction-mapping framework, the project will employ generative AI, graph neural networks, and multi-layer biological embeddings to model how genetically influenced pathways interface with host–microbiome metabolic networks. A particular focus will be the reconstruction of biosynthetic routes for anti-inflammatory metabolites—such as butyrate, indole-3-propionic acid, urolithin A, and bile-acid derivatives—and their linkage to relevant genetic variants, eQTLs, and microbiome-associated loci (mbQTLs).

Integrating causal genetic insights with mechanistic metabolic models, the project will prioritise:

1. Genetically validated therapeutic targets;
2. Drug repurposing opportunities supported by human genetic evidence;
3. Genetically informed dietary or microbiome-based interventions.

Primarily computational in scope, this project will deliver a genetically grounded, AI-enabled framework for elucidating host–microbiome mechanisms underlying LGSCI and identifying new therapeutic strategies for inflammation-driven disease.

Project 20

Cross-Ancestry Genomic and AI Approaches to Uncover Immune Phenotype Pathways in Autoimmune Disease

Supervisor 1: Dr David Morris, Medical and Molecular Genetics Faculty of Life Sciences & Medicine King's College London School of Medicine
Supervisor 2: Prof Timothy Vyse, Medical and Molecular Genetics Faculty of Life Sciences & Medicine King's College London School of Medicine

Contact: david.l.morris@kcl.ac.uk and timothy.vyse@kcl.ac.uk

Project Abstract: Autoimmune diseases such as systemic lupus erythematosus (SLE) represent a major challenge in precision medicine due to their complex genetic architectures and variable clinical manifestations across populations. While genome-wide association studies (GWAS) have identified numerous loci associated with SLE, understanding how these genetic variants influence immune system function and drive disease susceptibility remains an open question.
This project aims to integrate large-scale human genetic data with intermediate immune phenotypes such as differential white blood cell counts, immunoglobulin levels, and related immune biomarkers to identify and characterise shared genetic mechanisms across autoimmune conditions. Leveraging GWAS summary statistics from European, East Asian, African American, and Hispanic populations, we will apply statistical inference and AI-driven methods to (1) dissect the genetic architecture of immune phenotypes, (2) quantify genetic correlations and causal links with autoimmune diseases, and (3) identify population-specific and cross-ancestry risk variants that inform disease mechanisms.

Using approaches such as polygenic risk modelling, fine-mapping, and network-based integration of multi-omics data, the project will pinpoint molecular pathways and therapeutic targets that underlie immune dysregulation. The inclusion of diverse ancestry datasets enhances both discovery power and translational relevance, that will lead to findings that contribute to globally applicable insights in autoimmune disease genetics.

This work will provide a data-driven framework for linking genetic variation to immune system function and disease pathology, advancing target discovery and drug repurposing strategies. The project aligns closely with the programme’s goals by combining human genetics, systems biology, and AI to accelerate therapeutic innovation in immune-mediated disorders.

Project 21

Network inference, Validation, and In Silico Target Prediction for Metabolic Disorders and Ageing from Single-Cell Transcriptomics

Supervisor 1: Dr QueeLim Ch'ng, Centre for Developmental Neurobiology, School of Neuroscience, IoPPN
Supervisor 2: Dr Fursham Hamid, Centre for Developmental Neurobiology, School of Neuroscience, IoPPN

Contact: queelim@kcl.ac.uk and fursham.hamid@kcl.ac.uk

Project Abstract: Nutrient sensors link diet to metabolism and longevity, making them potential intervention points in obesity, metabolic diseases, and ageing. Nonetheless, leveraging known nutrient sensing pathways to identify therapeutic targets remains challenging because these pathways do not operate linearly. Instead, nutrient sensors are interconnected and embedded in larger neuroendocrine gene networks whose connectivity remains poorly understood. This knowledge gap represents a major rate-limiting step in target prediction and identification to tackle metabolic disorders and ageing.

To address this knowledge gap, we will exploit the roundworm C. elegans where every cell has been catalogued and where wild-type transcriptomic data is available for every cell. This unique resource permits a systems analysis of inter- and intracellular gene networks connecting highly conserved nutrient sensors with single-cell resolution for the whole organism.

This project will (1) computationally infer the connectivity patterns of highly conserved nutrient sensors in neuroendocrine gene networks from published single-cell transcriptomic data and validate network connectivity with single-nuclear RNA-Seq in nutrient sensing mutants; (2) map human genes and gene regulatory interactions involved in metabolism and ageing onto the network; (3) use the validated network to make in silico phenotypic predictions of conserved genes that exert the greatest effect on metabolism and ageing weighted by known roles in human metabolism and ageing, and test these predictions in loss-of-function mutants.

By leveraging the experimental ease of C. elegans and extensive human genetic datasets, the student will combine experimental transcriptomics and molecular genetics with diverse computational approaches including bioinformatics, machine learning/AI, network analysis, and image processing. Using predictive computational models to select and interpret experiments will reveal conserved genes and gene regulatory mechanisms that are the most promising therapeutic targets for interventions in metabolic diseases and ageing.

Project 22

Unlocking the therapeutic potentials of X chromosome inactivation escapees in sex-biased neurodevelopmental disorders

Supervisor 1: Dr Setsuko Sahara, Centre for Developmental Neurobiology, IoPPN
Supervisor 2: Prof Deepak P. Srivastava, BCN, SON, IoPPN

Contact: setsuko.sahara@kcl.ac.uk and Deepak.srivastava@kcl.ac.uk

Project Abstract: Females are generally more resilient to many diseases throughout life. This disparity is especially striking in developmental and neurodevelopmental disorders (NDDs), yet the biological basis of this early female protection remains largely unknown. How might female embryos establish greater resilience than males before hormonal influences emerge?

Biologically, sex is determined by the composition of the sex chromosomes—XX in females and XY in males. Classically, sex differences have been attributed mainly to gonadal hormones. The extra X chromosome in females was long considered neutral because an epigenetic mechanism known as X chromosome inactivation (XCI) equalises X-linked gene expression between the sexes. However, recent studies have revealed that a subset of X-linked genes, termed escapees, remain transcriptionally active or become reactivated. Elevated expression of such XCI escapees could generate transient or cell-type-specific sex biases during critical developmental windows, thereby conferring female neuroprotection against NDDs.

This project employs an integrative functional genomics and computational genetics strategy to translate XCI biology into disease mechanisms and therapeutic opportunities. Firstly, we will perform single-cell transcriptomic and epigenomic analyses of publicly available human developmental brain datasets to define a comprehensive “XCI Escapeome,” resolving cell-type- and developmental stage–specific escape patterns. Secondly, we will link escapees to NDD risk by integrating large-scale human genetic resources. Using disease GWAS, transcriptome-wide association studies, partitioned heritability, and colocalization analyses, we will identify escape genes whose genetically regulated dosage is causally implicated in autism and schizophrenia. Cell-type specific allele-specific expression analyses will provide orthogonal mechanistic support. Finally, we will prioritise high-confidence escapees for functional validation and drug repurposing. In male isogenic iPSC-derived neurons representing a dosage-vulnerable background, we will restore physiological escapee expression using tunable CRISPR activation and assess rescue of NDD-relevant cellular phenotypes. Together, this work integrates single-cell multi-omics, causal inference, and functional genomics to uncover X-linked protective mechanisms and deliver genetics-driven therapeutic targets for neurodevelopmental disorders.