Project title:

Transcriptomic stratification for improved Motor Neuron Disease modelling in human iPSCs

Supervision:

Primary Supervisor: Prof Kevin Talbot (Head of Department and Professor of Motor Neuron Biology) and Martina Hallegger (IMCM Fellow)
Day-to-Day Supervisor (Computational): Avigail Taylor (IMCM Technical Lead in Bioinformatics)
Day-to-Day Supervisor (Phenotypic analysis): Ruxandra Dafinca (will provide expertise on iPSC-derived neurons and phenotypic changes in MND cell lines)
GSK Supervisor: TBD (Andrew Goldfine as interim)

Summary: We recently showed that the RNA binding protein TDP-43, a key player in the neurodegenerative disease spectrum of FTD-ALS (Frontotemporal dementia, Amyotrophic lateral sclerosis), can assemble into condensates when bound to RNA and that this influences its RNA processing. The neuronal function of these condensates is still not completely understood on a molecular level or how they are regulated, become dysregulated, and, critically, how they impact RNA metabolism in diseases.

This DPhil project addresses a major open question in the field of what transcriptomic changes precede phenotypic alteration in iPSC-derived neuronal models of ALS/MND. The interplay between different RNAseq approaches will inform novel hypotheses to test by AI-ML, *in vitro* and then to experimentally verify *in cellulo* and patient data. The goal of this proposal is to define time points for compound treatment and to optimise the design of therapeutic interventions for ALS/MND.

We will develop machine-learning approaches to unravel aspects of neuronal RNA processing and thereby accelerate the translatability of transcriptomic and phenotypic data. We envision that AI-ML analysis of our unique transcriptomics datasets would contribute insights and tools to pave the way to new treatment approaches for ALS/MND.

Aims and Objectives: The DPhil student will analyse gene expression, alternative splicing and polyadenylation changes in MND neuronal models generated by the Talbot, Dafinca and Hallegger labs. Many of these data sets already exist, but all labs will add more data sets in the coming years.

To identify *bonafide* TDP-43 targets, these RNAseq data sets will be integrated with iCLIP data generated from these cell lines. iCLIP is a unique transcriptomic method that allows transcriptome-wide mapping of binding sites for RNA-binding proteins like TDP-43, which is central to MND pathology. TDP-43 RNA binding (iCLIP data) will be connected to RNA processing changes as seen in

RNAseq data sets (mRNAseq, 3'Seq); thereby direct RNA targets of TDP-43 are identified. All datasets will be uploaded to TERRA. Alternative splicing, polyadenylation and gene expression data will be integrated with iCLIP data.

We anticipate that disease-associated changes in the RNA binding behaviour of TDP-43 will lead to RNA network regulation changes, particularly early in the disease and would profoundly affect neuronal health. Such identified transcripts would represent novel therapeutic targets.

Alignment with therapeutic area and key scientific theme(s):

This project is closely aligned with Martina Hallegger's IMCM fellowship and will tie into proteomics analysis of ALS/MND patients' cerebrospinal fluid and blood samples conducted by the Talbot Team.

Project delivery:

Working in the cloud-based Oxford-GSK trusted research environment (TRE), the student will run bioinformatics pipelines for alternative splicing, polyadenylation and iCLIP analysis. These pipelines are instrumental to Martina Hallegger's fellowship project and the IMCM computational team will put these analysis pipelines in place in the first half of 2025, and this will allow us to test these ahead of the student's project start. The iCLIP pipeline is a publicly available containerised RNAseq analysis platform that the data team have agreed to implement on the Oxford-GSK TRE.

Martina Hallegger and Avigail Taylor (IMCM Technical Lead in Bioinformatics) will provide the overall and day-to-day supervision on the data analysis. The student will be embedded in the IMCM bioinformatics team, where they will receive the necessary training in bioinformatics, transcriptomics analysis and machine learning. The IMCM bioinformatics team hosts six full-time researchers with extensive experience of computational, statistical and ML analysis of massive 'omic data. Should the need arise, we will build collaborations with the Big Data Institute with world-leading ML expertise. The RNAseq data, including already published ones, will be uploaded to the Oxford-GSK TRE in the first half of 2025. The student will then analyse alternative splicing and polyadenylation changes in MND neuronal models generated by the Talbot, Dafinca and Hallegger labs. Using the above-mentioned bioinformatics pipelines, these data sets will be analysed for differential gene expression using DESeq2, or similar. Alternative splicing changes will be processed with, for example, MAJIQ and alternative polyadenylation changes will be quantified using DRIMSeq, or similar. Additional bioinformatics approaches will be used to quality control and interpret results. To identify *bonafide* TDP-43 targets iCLIP data will be integrated with RNAseq data sets (mRNAseq, 3'Seq). Prof Talbot, Dr Dafinca and Dr Hallegger will provide insights into the phenotypic changes observed in their iPSC derived motor neurons and transcriptomic changes will be linked to the onset of these phenotypes to identify mRNA targets with therapeutic potential.

Timelines: Firstly, the student will analyse available transcriptomic data from the Talbot, Dafinca and Hallegger labs on the Oxford-GSK TRE. In the first year the DPhil student will also develop machine

learning approaches for iCLIP (doi: 10.1186/s13059-023-03015-7; doi: 10.1093/bib/bbad307). They will train on datasets generated in Hek293 cells (Hallegger et al; doi: 10.1016/j.cell.2021.07.018) to define TDP-43 binding specificity. They will predict TDP-43 binding sites in neuronal transcripts based on RNAseq data from neuronal cells and post-mortem tissue generated by our team.

In years 2 and 3 iCLIP data generated in the Hallegger lab will be analysed from neurons derived from iPSC and from post-mortem neuronal tissue. Additional ML approaches will be used to assign iCLIP data to aged and diseased neuronal cell transcriptome to search for altered RNA binding behaviour in MND tissue.

The work ties in with the ongoing IMCM ALS project on biomarkers in banked blood / CSF samples. Gene expression changes and alternative splicing events leading to changes in peptides that can be detected in serum/CSF will be evaluated for use as diagnostic biomarkers.

Research environment:

The IMCM brings together the very best scientific, clinical, technological and computational expertise from Oxford University and GSK to form a unique industry/academic partnership. The Institute will develop disease-agnostic platforms to change the clinical practice of pathology, helping to identify and validate early potential drug targets and biomarkers to predict disease progression. The Institute is built around the IMCM Fellows and Oxford-GSK project teams, focusing on neurodegeneration. It provides an inclusive environment with a focus on positive and proactive mentorship, teamwork, and scientific creativity. The above-mentioned bioinformatics pipelines are instrumental to Martina Hallegger's fellowship project and will be implemented in the first half of 2025 by the IMCM Bioinformatics Team. Martina Hallegger and Avigail Taylor will provide the over-all and day-to-day supervision for the DPhil student, together with the IMCM bioinformatics team of six full-time researchers with extensive computational, statistical and ML analysis expertise. Prof Talbot and Dr Dafinca will provide their expertise on the phenotypic alterations observed in the iPSC derived neuronal models.