Appendix 1: Project Proposal

Guidance notes in italics. Limit proposal to two pages, excluding references.

Project title:

Transcriptomic stratification for improved Motor Neuron Disease modelling in human iPSCs

Supervision:

List proposed supervisors which must include representation from both Oxford and GSK. Indicate who will be the primary supervisor and provide day-to-day supervision of the research. If the GSK supervisor has not yet been agreed, please identify main contact.

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: Andrew Goldfine (TBC)

Summary:

Clearly describe the proposed work in terms of:

- *context*
- *the challenges the project addresses including main hypotheses*
- *aims and objective*
- *potential applications and benefits*
- *status of ethical approvals for data sharing with industry partners*

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 the RNA processing. The neuronal function of these condensates is still not completely understood on a molecular level or how they are regulated and 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 and Hallegger lab. Many of these data sets already exist, but both 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 TRE and integrated with alternative splicing and polyadenylation, gene expression and 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 present us with novel therapeutic target.

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

Identify which scientific theme(s) the proposal aligns. If alignment is tangential, provide additional justification.

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

Project delivery:

Explain how the project is designed so that it:

- *uses appropriate experimental methodology to address objectives*
- *facilitates the fellow's training in research methods related to data science*
- *can be delivered within the funding timeframe and has clear milestones against which to assess progress*
- *leverages existing Oxford-GSK infrastructure or builds new capabilities*

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 by the end of 2025, and this will allow us to test all pipelines ahead of the start of the student's project. 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 appropriate training in bioinformatics, transcriptomics analysis and machine learning. The student will analyse alternative splicing and polyadenylation changes in MND neuronal models generated by the Talbot and Hallegger lab, the RNAseq data, including already published ones, will be uploaded to the Oxford-GSK TRE in the first half of 2026.

Using the above-mentioned bioinformatics pipelines, these data 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 results and interpret results. To identify bonafide TDP-43 targets all RNAseq data (mRNAseq, 3'Seq) will be integrated with iCLIP data.

Timelines: Firstly, the student will analyse available transcriptomic data from the Talbot and Hallegger lab through TRE. In the first year of the DPhil project the 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 the Talbot and Hallegger lab. Transcripts that are expressed in both neurons and Hek293 lines will act as an internal positive control of the quality of prediction.

In year 2 and 3 iCLIP data generated in the Hallegger lab will be analysed from neurons derived from iPSC from the Talbot, Hallegger lab 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:

Provide additional information on support available for the fellow including:

- *brief details on whether the research group or institution are currently using similar methodologies or exploring similar hypotheses*
- *any additional financial or practical support specific to the research group or institution*
- *the extent and frequency of any clinical commitments expected of the fellow*

Our team addresses the unique question of what transcriptomic changes are linked to the changes in TDP-43 assembly behaviour and will connect RNA interactome studies with RNAseq data to provide a better understanding of TDP-43 dysfunction in MND pathology. The Talbot and Dafinca Lab has extensive expertise in modelling MND in iPSC derived neuronal cells. The Hallegger lab has all the require transcriptomics expertise. Martina Hallegger is an expert in TDP-43 RNA interaction and TDP-43 granule formation and their link to transcriptomic changes. Avigail Taylor has ample expertise in implementing transcriptomics analysis pipelines and AI-ML projects. Together with the IMCM computational team we will develop a technical platform to address key question of transcriptomic signature of neurodegeneration which will allow us with the aim to identify and validate early potential drug targets and biomarkers to predict disease progression.

Institutional support: The CHG provides an excellent environment with iPSC facilities, excellent transcriptomics and bioinformatics extertise and an high-performance computing cluster for large-scale data analysis.

Clinical Commitments: The student will be able to obtain an honorary clinical contract if required, to support ongoing training but the main focus of the studentship will be on research.