Identifying the master regulators of transcription in multiple sclerosis by combining genetic analysis with single cell expression profiling

Sawcer & Stegle

Background and data:

Multiple sclerosis is an autoimmune disease of the central nervous system that results in chronic neurological disability for most affected individuals and is a major drain on the health economy of the United Kingdom (UK). Susceptibility to the disease is highly heritable and Genomewide Association Studies (GWAS) have been spectacularly successful at identifying relevant risk loci, with more than 200 already mapped. These associated variants overwhelmingly lie in regulatory regions of the genome that are active in immune cells, indicating that the disease most likely results from altered expression of otherwise normal genes in critically important immune cell sub-types. The aberrant lymphocytes driving the disease are concentrated in the cerebrospinal fluid (CSF) of affected individuals and therefore, supported by a grant from the UK multiple sclerosis society, we are generating single cell transcriptomes from human CSF lymphocytes using the 10X system. Ultimately we will process approximately 500 cells from each of 50 cases and 50 controls; some 50,000 transcriptomes in total. The 10X system uses unique molecular identifiers (UMI) to allow quantitative and well-calibrated transcripts counts to be deduced from next generation sequencing (NGS). These data generation is already ongoing and will form the basis of this proposed EBPOD project. Given the need for samples to be processed as rapidly as possible after collection the wet work for the project is being completed at the Cambridge Single Cell Analysis Clinical Core Facility in collaboration with Prof Bertie Gottgens. In addition to these newly generated single-cell profiles, we have also generated bulk RNA-seq data from multiple sclerosis (n=100) and systemic lupus erythematosus (n=100) patients and controls (n=100), which can also be leveraged in the analysis.

The EBPOD candidate will work collaboratively between the Stegle group at EBI and the Sawcer group at the BRC. A major question will be to understand disease-specific changes single-cell transcriptome profiles and the integration with genetic risk factors for CSF.


An initial aim will be to assess difference in single-cell transcriptome between cases and controls. Building on the extensive experience in single-cell RNA-seq analysis in the Stegle team, we will assess a broad range of scRNA-seq derived phenotypes, thereby assessing changes in gene expression level, gene expression heterogeneity between cells and disease-specific sub populations of cells. We will also consider these transcriptomic signatures for identifying subgroups of patients, thereby assessing their potential as predictive biomarkers for disease prognosis.
Identifying the master regulators of transcription in multiple sclerosis by combining genetic analysis with single cell expression profiling

**Aim 2:** Genetic analysis of single-cell transcriptomes, including disease risk.

Disease risk has a strong genetic component. The Stegle team has previously developed statistical approaches for tying together human genetic variation and single-cell readouts\(^2\,^3\), which we will here extend to the setting of case/control cohorts as required for this study. The Sawcer team has identified a large number of individual risk variants, which we will consider individually and for calculating polygenic risk scores in subjects. Interrogating these data with advanced statistical modelling, at the single-cell level, we will assess the effects of genetic risk factors on gene expression level, expression variance and changes in cell types and states.

**Aim 3:** Identification of disease-associated molecular networks and targets.

Finally, we seek to use Bioinformatics approaches for integrating the regulatory associations identified into disease-associated networks. We will use reverse engineering to identify the critical master regulators that underlie these associations\(^4\). This approach acknowledges that although the particular set of risk alleles underlying the development of a disease state are certain to differ between subjects they will ultimately exert their effects by disrupting key aspect of the transcriptional network.

**Partners and training opportunities**

The candidate will be embedded in a strong multidisciplinary team. The Stegle group at the EBI will provide access to cutting-edge statistical and computational expertise, uniquely combining the areas of statistical genetics and single-cell biology. Stephen Sawcer is the Professor of Neurological Genetics and has leading expertise in the genetics of multiple sclerosis and other neurological conditions for over twenty years. The fellow will be fully embedded in both groups and be able to acquire a profile at the interface of human genetics, single-cell genomics and disease biology.

**References:**