



Application number/Title: 43730 - Novel methods for the use of hierarchically-structured healthcare data in genetic association analyses

Applicant PI: Dr Henrik Edgren

Application Institution: MediSapuens Ltd, Helsinki, Finland

Keywords provided by the Applicant PI to describe the research project:
phewas, treewas, data-curation, genetics, ontologies

Application Lay Summary:

Over the last years, research on the UK Biobank data has demonstrated the power of large integrated data sets of genomic and health status data in identifying genetic contributions to many human diseases. A key factor has been the availability of extensive, high quality, clinical and other health related data on the genotyped UK Biobank samples. However, the detail at which clinical observations are recorded presents a trade-off: as recording of clinical observations gets more detailed, disease - variant associations become more specific. At the same time, however, the more detailed the observations are, the lower the number of samples with any given observation, decreasing statistical power to identify associations. Over the next 24 months, this project aims to develop analytical methods that use hierarchically encoded clinical observations, such as disease diagnoses, to increase the statistical power to identify genetic associations for these observations. In addition, we will compare data already encoded in UK Biobank using hierarchical terminologies to association results obtained on the same samples when data is encoded with ontologies, such as SNOMED CT and Disease Ontology. The aim being to understand the applicability of different ontologies and controlled vocabularies for specific disease areas. The research proposal outlined here ultimately aims to improve the identification of genetic associations for human health and disease related traits, thereby contributing to the development of better prevention strategies, medical diagnostics and treatments.