Project Details	
Project Code	PHS27Ex Wright
Title	Sequencing babies: understanding penetrance of childhood epilepsy to inform newborn screening
Research Theme	Population Health Sciences
Project Type	Dry lab
Summary	Many countries are now planning to sequence the complete DNA of babies at birth to screen for hundreds of genetic conditions. This PhD project will generate the evidence to inform which genes and variants should be included in newborn genome screening. Using large datasets from clinical and population cohorts, the student will analyse the prevalence and penetrance of variants associated with monogenic epilepsy and other neurodevelopmental disorders to expand our understanding of genotype-phenotype correlations in these conditions. This work will improve understanding of the risks associated with genetic variants, and guide which conditions should be included in newborn genome screening.
Description	BACKGROUND: All babies in the UK are now screened at birth for ten rare but treatable diseases through the newborn bloodspot test. However, many more genetic conditions exist that are very rare but also potentially treatable - many of which affect neurodevelopment -and could be detected simultaneously using whole genome sequencing. The NHS has recently introduced a pilot scheme, the Generation study, which aims to sequence the genomes of ~100,000 newborns to identify genetic variants associated with >200 rare diseases. But there is currently a gap in available evidence for interpreting genetic variants associated with disease, especially without the presence of associated features ascertained via clinical cohorts. The availability of large population cohorts - including whole genome sequencing data from millions of individuals - now makes it possible to investigate genetic causes of disease outside of a clinical context, which is crucial to implementing genome screening as part of a public health programme. As part of an ongoing MRC-funded project, Exeter's Genomics of Rare Disease team has done numerous projects generating evidence from population cohorts to inform which genes and variants should be included in newborn genome sequencing (see for example, numerous papers in https://www.nature.com/collections/djigcbgbif). However, further many more conditions remain to be investigated, and methodological development is needed to evaluate related conditions simultaneously. AIMS: This PhD project will use large population and clinical datasets to investigate the prevalence and penetrance of variants associated with childhood-onset epilepsies and neurodevelopmental disorders to expand our understanding in genotype-phenotype correlations of these conditions. This work will provide a greater understanding of the risk associated with specific genetic variants, and ultimately help create guidance on which genes should be included in newborn screening. OBJECTIVES: 1. Survey which genes/variants are associated with childhood-onset epilepsies and neurodevelopmental disorders and might meet newborn screening criteria (as laid out by the Generation Study).

	2. Curate disease-associated variants and disease mechanisms 3. Use UK Biobank (~500,000 adults) and All of Us (~450,000 adults) cohorts to look for disease-associated variants from genome sequencing data and investigate disease-relevant phenotypes. This will include extensive curation of genetic variants and health records, as well as statistical testing of associated biomarkers (e.g. transcriptomic and proteomic data) and related traits (e.g. fluid intelligence, employment, etc). 4. Evaluate the effect of relevant polygenic risk scores and other modifiers on penetrance. 5. Use Our Future Health (~5 million adults) to validate findings using imputed data from genotyping arrays. 6. Use locally-available clinical datasets to compare disease-associated variant types and treatment outcomes. 7. Investigate the use of AI-methodology to enable evaluation of multiple related genetic conditions simultaneously. OWNERSHIP: Although the project will start with epilepsies and neurodevelopmental disorders, there is plenty of scope for the student to focus more deeply on specific conditions within this remit, or to expand out to include related conditions under consideration for newborn genome screening. Moreover, the UK Biobank and All of Us cohorts contain not only whole genome sequence data linked with electronic health records but also numerous biomarkers (transcriptomic, proteomic, metabolomic) which can be evaluated to provide additional evidence. The student will also be empowered to steer the project towards using AI-methodologies to investigate multiple conditions simultaneously. TECHNIQUES/TRAINING: The student will need to use a variety of techniques to fulfill the research aims, including bioinformatics, genetic variant interpretation, population genetics, statistical genetics, and population health sciences. Embedded in the wider genomics team at Exeter, the student will be well supported by a network of genomic scientists with extensive experience working with large genomic datasets.
Can be completed part time?	Yes
Supervisory Team	
Lead Supervisor	
Name	Professor Caroline Wright
Affiliation	Exeter
College/Faculty	Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Email Address	caroline.wright@exeter.ac.uk
Co-Supervisor 1	
Name	Dr Leigh Jackson
Affiliation	Exeter
College/Faculty	Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Co-Supervisor 2	
Name	Professor Nick Timpson

Affiliation	Bristol
College/Faculty	Medical School
Department/School	Population Health Sciences
Co-Supervisor 3	
Name	Dr James Fasham
Affiliation	Exeter
College/Faculty	Health and Life Sciences
Department/School	Clinical and Biomedical Sciences