## 1.1 Clinical Genetics

There are over 7000 recognised Mendelian diseases (1). Whilst individually rare, approximately 8% of live births are reported to have a genetic disorder by early adulthood (2). Yet, it is estimated that half of patients with a suspected genetic condition fail to receive a molecular diagnosis (3,4). The term ‘diagnostic odyssey’ is used to describe the period of time from initial disease onset to final diagnosis and can span several years. The absence of a clinical diagnosis can mean missed therapeutic opportunities and the inability to provide accurate information on prognosis and recurrence risk, which can result in anxiety for both patients and their families. There is also a substantial economic burden associated with continued testing.

In medical genetics, the differential diagnosis process is complicated by the vast number of potential disorders, many of which have heterogenic and unspecific symptoms of varying expression and penetrance. Moreover, a causative gene has not yet been identified for approximately half the Mendelian diseases. For disorders which do have associated disease genes, the possibility of rarer molecular associations still remains. The current trajectory of disease gene discovery is reported as 263 novel findings per year (5).

## 1.2 Hypertrophic Cardiomyopathy (HCM)

### 1.2.1 Disease Background

Hypertrophic Cardiomyopathy (HCM) is the most common inherited cardiac disorder. It is the leading cause of sudden death in young people (< 35) and athletes, and is a major cause of heart failure for all ages (6). The disease can be diagnosed at any age but most frequently presents in the third decade of life (7). Females tend to present with HCM later in life than men but are more symptomatic (8). However, HCM-related mortality is equal for males and females and a distinct global patterning has not been recorded.

Clinically, HCM is defined by thickening of the heart muscle that cannot be attributed to secondary causes. More specifically, left ventricular hypertrophy (LVH) measuring greater than 14mm is recommended by the European Society of Cardiologists as a diagnostic criterion. The average ventricle size of a normal adult male is 10mm (9). This thickening causes the heart muscle to become stiff and consequentially reduces the efficiency of blood flow around the body.

### 1.2.2 Clinical Diagnosis of HCM

Providing a diagnosis of HCM based solely on clinical features is complicated by the heterogeneity of phenotypes, which show high inter- and intra- family variability and can appear at any age. Familial disease-causing variants also display incomplete penetrance, i.e. known carriers may remain asymptomatic for their entire lives (10). Moreover, the presentation of LVH is not limited to HCM and can occur as a result of numerous other conditions such as mitochondrial disorders or glycogen storage diseases (11,12). Such conditions are distinct from familial HCM and require different clinical management. It is therefore important to understand the genetic basis of HCM.

### 1.2.3 Genetics of HCM

Phenotypically, HCM affects approximately 1:500 people worldwide. However, it is predicted that the genetic prevalence could be as high as 1:200 (6). HCM most frequently presents as a single gene disorder with an autosomal dominant pattern of inheritance. Approximately 60% of patient cases have a recognised family history (13).

Familial HCM is predominantly caused by pathogenic variants in sarcomere-related genes that encode the contractile machinery of the heart… expand.

### 1.2.4 Clinical Genetic Testing for HCM

Clinical genetic testing for inherited HCM has been available for over 15 years and makes up a key aspect of patient care. As sequencing technology has advanced, this has progressed from single gene analysis to gene panels of well-established HCM genes (14). Panel testing is widely available in the UK as an NHS service and the targeted sequencing of most genes included on current panels is now recommended by clinical guidelines (9).

As well as providing a definitive diagnosis for the patient, genetic testing is important for the identification of at-risk family members so they can either be clinically managed or discharged [8]. However, due to the clinical and genetic heterogeneity of HCM, a pathogenic variant is only found in approximately 30-40% of cases [9, 10].

In part, this is because variant interpretation in HCM is challenging. Segregation studies are complicated by variable penetrance (15). In addition, when a novel variant is identified, there is often limited functional information available about the domain in which it resides. Molecular diagnosis is further complicated when the condition is not Mendelian, but instead occurs as a result of variants in a combination of predisposing genes and contributing environmental factors (10). Finally, patients can also harbour a variant in an unknown gene or gene that is not included in the gene panel. Determining whether or not HCM in these patients has a genetic basis is a key research priority.

## 1.3 Exome and Genome Sequencing

### 1.3.1 The 100 000 Genomes Project

Exome and genome sequencing have been hugely successful in furthering disease gene discovery and are starting to be used in clinical diagnostics (16,17). In 2012, Genomics England launched the 100,000 Genomes Project in collaboration with NHS England. The aim of the project was to sequence 100,000 genomes from 70,000 individuals with rare diseases and their families as well as patients with cancer. As of the 1st October 2018, 87,231 genomes had been sequenced. Ultimately, the purpose of the project is to provide a new genomic medicine service for the NHS and to introduce genome sequencing into clinical practice. By combining genome sequence data with medical records, it is hoped that this large-scale resource will provide a better understanding of a ‘normal’ reference genome, facilitate disease-gene discovery, improve diagnostics and further medical research (18). HCM is one of the disorders included as part of the rare disease arm of the 100,000 Genomes Project. However, the clinical utility of genome sequencing in autosomal dominant clinically heterogeneous cases such as HCM remains unclear (19).

### 1.3.2 Challenges of Variant Interpretation

The analysis of exome and genome sequence data is still a major challenge. Studies suggest that an average exome harbours more than 30,000 variants compared with the reference sequence. Of these, approximately 10,000 are predicted to give rise to non-synonymous amino acid changes, small insertions or deletions or alterations of conserved splice site residues (20). Many methods are to filter variants for potential pathogenicity, such as: (1) frequency in general and disease populations (2) conservation and (3) the chemical (dis)similarity between the wild type and variant amino acid. However, a standard human genome is believed to encompass approximately 100 loss-of-function variants and 20 inactivated genes (21).

Further, recent large scale studies comparing HCM and control cohorts have identified previously considered well-established pathogenic HCM variants at population frequencies that are incompatible with pathogenicity. This is of particular importance when considering the possibility and implications of misdiagnosis (22).

Collectively, this suggests that merely filtering variants based on rarity and predicted pathogenicity may not be the best criterions for identifying candidate disease genes in exome and genome sequencing data.

## 1.3 Exomiser

A number of computational tools have been developed to integrate complex data sets and deliver a ranked list of variants to offer an indication of those that warrant further investigation. One such tool is the Exomiser (23). The Exomiser tool calculates both a variant and gene score before combining them using a logistical regression model to generate a final score used for ranking. The variant score is based on pedigree analysis, predicted pathogenicity and the rarity of the variant in the 1000 Genome Project and Exome Server Project datasets. The gene score is calculated via a method chosen by the user. These additional algorithms include the analysis of protein-protein interaction networks and phenotype comparisons between the patient case and both curated human disease databases and a number of model organisms . Each is detailed below.

### 1.3.1 PHIVE

(all exomiser algorithms, include a time line and try to discuss HPO terms, ontology, model organisms and monarch initiative)

GENOMISER

AIMS AND OBJECTIVES