UNIVERSITY OF STIRLING

An increase in copy number of Myosin Light Chain Kinase 1 associates with increased force production in Lithuanian athletes

By

David John Hunter BSc (Hons)

Supervisors:

Dr Colin Moran

Dr Iain Gallagher

Research Thesis for the award of M.Phil

School of Sport

University of Stirling.

August 2014

Declaration

I declare that this thesis and the work presented in it are my own and have been generated by me as the result of my research. It has not been submitted anywhere else for any award. Where other sources of information have been used they have been clearly acknowledged and referenced.

Acknowledgements

I would like to thank all of the academic staff from the University of Stirling HESRG who helped me throughout my studies and made the last year an enjoyable experience. A special thanks goes to Colin Moran for his supervision during this thesis. His guidance, support and advice throughout has not only allowed me to complete this research but also provided me with valuable experience for future. Thanks also go to all of the individuals involved in the collection of the GELA cohort.

Further thanks go to all my family, friends and fellow postgrads that have supported academically and socially throughout my M.Phil. Without you all, this last year would not have been as enjoyable as it has proven to be.

I would like to dedicate this thesis to my mum who made me the man I am today and who I miss every day.

Table of contents

Declaration ii
Acknowledgements iii
Table of contents iv
List of Tables vi
List of Figures vii
Abstract 1
Abbreviations 2
Chapter 1 - Introduction 4
1.1 Background 4
1.2 Heritability of traits 7
1.3 Genetic variation 9
1.4 Structural Variants 11
1.5 Copy number variation 12
1.6 Myosin Light Chain Kinase 16
1.7 Aims & Objectives 21
Chapter 2 - Materials and Methods 22
2.1 Ethical Approval 22
2.2 Participant characteristics 22
2.3Phenotype data 24
2.4 DNA Extraction 25
2.5 Copy Number Assays 25
2.6 Copy number determination 27
2.7 Statistical Analysis 29
2.7.1 Comparison of copy number / genotype between groups 29
2.7.2Association of copy number with physiological variables 30
Chapter 3 - Results 31
3.1 Copy number detection in each group 31
3.2 Copy number differences between groups 32
3.3 Copy number associations with quantitative measurements 33
3.4 Subgroup copy number associations with quantitative variables 37
Chapter 4 - Discussion 41
4.1 Main Findings 41

4.2 MYLK1 41
4.2.1 Copy number and athlete status 41
4.2.2 Level of Competition 44
4.2.3 Quantitative data 46
4.2.4 Mechanisms 49
4.3 MYLK-AS1 50
4.4 Summary 52
4.4.1 Combined findings 52
4.4.2 Strengths 55
4.4.3 Limitations 55
4.4.4 Future work 56
Chapter 5 - Conclusion 57
References 58

Appendices:

Appendix A – The total number of athletes and number of athletes within each level of competition who perform each of the sports. 67
Appendix B – Details on the physiological tests selected for phenotype determination 68
Apendix C – Concentrations of the DNA samples used in the present study 70
Appendix D - Pearson's correlation coefficient results between END for MYLK1 copy number and quantitative variables. IKR; isokinetic right, IKL; isokinetic left 73
Appendix E – Increases in performance per copy number in the whole cohort for quantitative phenotypes which displayed significant correlations after BH corrections. IKR: isokinetic right, IKL; isokinetic left. 74
Appendix F – Increases in performance per copy number in STP for quantitative phenotypes which displayed significant correlations after BH corrections. 75

List of Tables

Table 1 - Problems that can develop through the use of a single gene for the relative quantification of structural variants. 12

Table 2 - Descriptive characteristics of the groups of participants. Values are means ± 95% CI. FFM; fat free mass, STP; strength and power athletes, END; endurance athletes. 24

Table 3 – The copy number of the individuals within each of the groups. Values are the total number with the percentage of the group in brackets. CN; copy number, STP; strength and power athlete, END; endurance athlete 32

Table 4 - Odds Ratios for athlete group vs copy number for MYLK1. Odds Ratio of N/A is because none of the controls have a copy number of one for MYLK1. STP; strength and power athletes, CI; Confidence interval 33

Table 5 - Pearson's correlation coefficient results for MYLK1 copy number and quantitative variables before and after correction for subgroup. Significant ($p<0.05$) values are highlighted in bold. Values marked by * have a p-value lower than the acceptable alpha value according to the BH correction. kgFFM; kilogram of fat free mass, IKR, isokinetic right; IKL, isokinetic left 35

Table 6 - Pearson's correlation coefficient results for MYLK1-AS1 copy number and quantitative variables before and after correction for subgroup. Significant ($p<0.05$) values are highlighted in bold. Values marked by * have a p-value lower than the acceptable alpha value according to the BH correction. kgFFM; kilogram of fat free mass IKR, isokinetic right; IKL, isokinetic left 36

Table 7 - Pearson's correlation coefficient results between STP for MYLK1 copy number and quantitative variables. Correlations with $p<0.05$ are highlighted in bold and values marked by * have a p-value lower than the acceptable alpha value according to the BH correction. kgFFM; kilogram of fat free mass IKR, isokinetic right; IKL, isokinetic left 38

List of Figures

Figure 1 - Role of MLCK in the contraction of skeletal and smooth muscle (adapted from (Takashima, 2009)). SR; sarcoplasmic reticulum, Ca2+; calcium ions, MLCK; myosin light chain kinase, sk; skeletal muscle, sm; smooth muscle, RLC; regulatory light chain. 18

Figure 2 – Location of the MYLK1 and MYLK-AS1 sequences on Chromosome 3 according to the Genome Reference Consortium Human Build 38. Bp; base pair location on chromosome 3, MYLK1; Myosin light chain kinase 1, MYLK-AS1; Myosin light chain kinase-antisense 1. 27

Figure 3 – Mean peak force produced (N*m) for isokinetic right arm extension of the elbow at 180 °/s for each copy number in the whole cohort and the STP alone. Copy numbers which do not share the same letter are significantly different from each other with each group. CN;copy number, IKR Ext; isokinetic right arm Extension, STP; strength and power athletes. 39

Figure 4 - Mean static jump height (cm) for each copy number in the whole cohort and the STP. Copy numbers which do not share the same letter are significantly different from each other with each group. CN;copy number, STP; strength and power athletes. 40

Figure 5 - Mean 30 m sprint time (presented as s/kgFFM) as for each copy number in the whole cohort and the STP alone. Copy numbers which do not share the same letter are significantly different from each other with each group. CN;copy number,kgFFM; kilograms of fat free mass, STP; strength and power athletes.....40

Abstract

Copy number variation has been linked to the development of various diseases, however copy number variants have not been investigated for associations with athletic performance. The present study investigated copy number variation of the Myosin Light Chain Kinase (MYLK)1 and MYLK-Antisense 1 (MYLK-AS1) genes for associations with athletic performance in a cohort of Lithuanian athletes and controls. We hypothesised that increased MYLK1 copy number and decreased MYLK-AS1 copy number would be associated with strength and power athletes (STP) and measures of muscle performance. DNA was extracted from blood samples of 407 athletes and controls. Copy number of the target genes was determined using a multiplexed quantitative-polymerase chain reaction and the use of a multicopy reference assay. A higher MYLK1 copy number was overrepresented in the STP compared with controls ($p=0.028$[OR=9.97, 95% CI: 1.19-83.81); however endurance athletes did not differ from controls or STP ($p \geq 0.661$; $\chi^2 \leq 0.829$). Positive correlations between MYLK1 copy number and strength and power performance were detected. Individuals with three copies were able to produce more power in isokinetic tests, jump higher and sprint faster than individuals with two copies in the whole cohort and these correlations were stronger in the STP. The strongest correlations found were with MYLK copy number and isokinetic flexion of the arm in STP ($p=0.003$, $R^2=7.78$). MYLK-AS1 did not associate with athlete status or measures of athletic performance ($p>0.05$). The improvements in performance with increased copy number indicate that copy number variants are associated with improved skeletal muscle phenotypes. These results, further current knowledge of how genetic variants underpin muscle phenotypes and indicate that MYLK1 potentially could be used as a target for improvements in athletic performance and treatment of muscle wasting disorders.

Abbreviations

Deoxyribonucleic acid	DNA
Ribonucleic acid	RNA
Monozygotic	MZ
Dizygotic	DZ
Single nucleotide polymorphism	SNP
α-Actinin-3	ACTN3
Structural variant	SV
Polymerase chain reaction	PCR
Fluorescence in situ hybridization	FISH
Quantitative PCR	qPCR
Copy number variant	CNV
Salivary amylase	AMY
Survival of motor neuron	SMN
Myosin light chain kinase (gene)	MYLK
Regulatory light chain of myosin	RLC
Myosin light chain kinase (protein)	MLCK
Smooth muscle MLCK	smMLCK

Skeletal muscle MLCK	skMLCK
Cardiac muscle MLCK	cmMLCK
MYLK-antisense1	MYLK-AS1
Strength and power athlete	STP
Endurance athlete	END
Crossing point	Cp
Odds Ratio	OR
Benjamini-Hochberg	BH
Post-activation Potentiation	PAP

Chapter 1 - Introduction

1.1 Background

Skeletal muscle is the largest organ in the human body making up approximately 30% of body mass (Kostek *et al.*, 2010; Pescatello *et al.*, 2013). The amount of force that can be generated (i.e. muscle strength), is the primary measure of skeletal muscle function and is often used as an indicator of health. The functional capacity of skeletal muscle is of critical importance because the maintenance of strength is essential in the prevention and treatment of muscle wasting diseases, such as sarcopenia (Roth *et al.*, 2000; Roubenoff, 2000) and metabolic disorders, such as type II Diabetes (Boulé *et al.*, 2001). Skeletal muscle underpins other processes such as locomotion and movement (Kostek *et al.*, 2010). Since skeletal muscle is critically important for force production and movement, it is clear that the quality of muscle is a regulator of both sporting performance and health. The age related decline in both muscle mass and strength has been termed sarcopenia (Roth *et al.*, 2000; Roubenoff, 2000; Metter *et al.*, 2002; Goodpaster *et al.*, 2006; Mitchell *et al.*, 2012). Studies have shown the steady loss of skeletal muscle begins at 40-50 years of age and the rate of loss is of such a degree that by the age of 80 the typical loss of skeletal muscle fibres from the limbs is 50% (Faulkner *et al.*, 2007; Mitchell *et al.*, 2012). Despite the maintenance of a high level of physical fitness, the performance of 'master' athletes typically declines after the age of 40 and by the age of 80 can be reduced by up to 65% (Faulkner *et al.*, 2007). However, 'master' athletes are stronger than master non-athletes of the same age (Hawkins *et al.*, 2003). Muscle strength is lost at a greater rate than muscle mass(Faulkner *et al.*, 2007; Mitchell *et al.*, 2012). Despite a strong correlation between muscle mass and strength, in older individuals gains in lean mass can occur with decrements in muscle strength, indicating that the quality of the muscle is the important factor (Goodpaster *et al.*, 2006; Mitchell *et al.*, 2012). Sarcopenia is

an important process to understand because it can lead to disability, falls and the loss of independence in otherwise healthy elderly individuals (Roubenoff, 2000; Metter *et al.*, 2002; Goodpaster *et al.*, 2006; Mitchell *et al.*, 2012). It has been shown that individuals with higher muscle strength, in large muscle groups, have a greater survival rate from a range of diseases compared to individuals with lower muscle strength (Metter *et al.*, 2002; Goodpaster *et al.*, 2006; Ruiz *et al.*, 2009). Thus, understanding the determinants of muscle strength is an important area of scientific research from both a health and performance perspective.

Muscle strength is a highly variable trait and at the two opposing ends of the muscle strength scale there are elite athletes and individuals who suffer from muscle wasting diseases. Various factors determine muscle strength including environmental factors, such as nutrition, level of physical activity (Tiainen *et al.*, 2009), and behavioural factors, such as motivation to train and level of pain tolerance (Sallis *et al.*, 1992; Boutelle *et al.*, 2004). Physiological factors also have been shown to influence strength through a myriad of different ways such as fibre composition, cross sectional area and size of the muscles (Kostek *et al.*, 2010). There is a significant genetic component to these behavioural and physiological factors. The interaction between the environment and the genetics of an individual are just as important in phenotype determination as either factor alone (Tiainen *et al.*, 2009). Individuals who possess advantageous genetics and control their environment correctly have the greatest chance of developing into elite athletes. Therefore, muscle strength is a biological process which is determined which is partially determined by an individual's genetics and supported by both the environment and the interaction between the genetics and the environment (Kostek *et al.*, 2010).

The human haploid genome is encoded on deoxyribonucleic acid (DNA) sequences which consist of approximately 3 billion nucleotides (Kruglyak & Nickerson, 2001). DNA encodes 20,687 protein coding genes in all individuals (Bernstein *et al.*, 2012). The sequence of any two human genomes are highly similar with estimates of over 99% identical, which leaves 3 million potential nucleotides that determine the large phenotypic differences both within and across different populations (Kruglyak & Nickerson, 2001). The variation in the sequence of DNA causes the genetic differences in phenotypes. If a genomic variant is found at greater than 1% within a population then it is termed a common variant, whereas if the variant is found at less than 1% then it is known as a rare variant. Typically rare variants have a large impact on the function of a gene and are consequently less tolerable; therefore reducing the likelihood of them being passed on to future generations. Common variants tend to be more tolerable and have small to no effect on a gene, thus giving them a higher chance to be passed on to future generations than mutations (Pollex & Hegele, 2007).

Less than 2% of the human genome encodes for protein coding genes (Bernstein *et al.*, 2012), and at least 5% of the genome has been shown to be under selection. Therefore it was thought regions other than protein coding regions were functional (Pheasant & Mattick, 2007). The remaining 95% of the genome was referred to as 'junk' DNA because it had no obvious functional role. The completion of ENCODE project discovered that the majority of the 'junk' DNA has some form of biochemical function. A total of 80% of the genome has been shown to have some form of functional role on either RNA or chromatin (Bernstein *et al.*, 2012). Most of this percentage acts post transcriptionally to modify coding RNAs to alter regulation or expression of coding RNA (Palazzo & Gregory, 2014). One such type of non-coding gene which acts to control the level of mRNA is an antisense gene. Antisense transcripts have been shown to regulate the expression of sense transcripts through RNA-

RNA interactions. These interactions can cause gene silencing through chromatin modifications, such as DNA methylation and histone modification or interference where the sense and antisense transcripts hybridize and are subsequently destroyed (Werner, 2013).

1.2 Heritability of traits

Heritability is defined as the proportion of the variability of a phenotype which is explained by genetic factors within a population (Visscher *et al.*, 2008). Twin and family studies are the two main study designs used to detect the heritable component of traits associated with athletic performance (Beunen & Thomis, 2006). Family studies examine the similarity between siblings and their parents to estimate the variation in the genetic component. These studies typically overestimate the genetic component due to conflation with shared environment. Twin studies examine monozygotic (MZ) and dizygotic (DZ) twins to determine the genetic and environmental impact of a trait. MZ twins have nearly identical genomes and DZ twins only share ~50% of their genetic material (Shih *et al.*, 2004). If the phenotypes of MZ twins are more similar then it is said there is a genetic component, whereas if the phenotype of DZ twins is equally similar then it is said there is a purely environmental impact on the heritability of the trait.

Twin and family studies only show the heritability of a trait within the specific population which is being examined because factors such as race, sex and age can all affect the heritability differently depending on the population (Visscher *et al.*, 2008). When the heritability of height is examined, the population specific aspect of heritability is highlighted. In Caucasian populations heritability of height in males has been shown to range from 87-93%, whereas in females the heritability is estimated to be 68-84% (Silventoinen *et al.*,

2003). The heritability of height also has been examined in a range of different populations of African descent. In Nigerians the heritability of height is estimated at 62%, whereas in African Americans the heritability is estimated to be 87%, while in Jamaicans the heritability is 76% (Luke *et al.*, 2001). The age of the individuals examined influences heritability. Typically the heritability of a trait will change over the course of a lifetime because of different environmental impacts at certain points in life. For example body mass at birth is dependent on the uterine environment, as an infant body mass is determined by the maternal milk production and as an adult there is no maternal component which influences body mass (Visscher *et al.*, 2008). Twin and family studies indicate there are several aspects of athletic performance which have been shown to have large heritable aspects associated with them.

The heritability of aerobic performance is one of the most studied athletic performance traits. The early studies on heritability generally over or under estimated the genetic component because of a lack of control for bias such as, age or gender (Bouchard *et al.*, 1986). The HERITAGE study is one of the recent studies which controlled for several factors (age, sex and body mass) and predicted the heritability of $\dot{V}O_2$ max and the trainability of $\dot{V}O_2$ max was approximately 50% in sedentary individuals (Bouchard *et al.*, 1998, 1999). Several other traits of athletic performance have been studied including muscle fibre composition, anaerobic performance and muscle strength and power. More recent twin studies have estimated that the genetic element of an individual's athletic performance is 31-85% (De Moor *et al.*, 2007).

The heritability of muscle strength traits is not as clear because in the literature there are a wide range of heritability estimates. Twin and family studies have shown there is a large

heritable factor associated with muscle strength; however, inconsistencies in study design have led to wide estimates in the heritability estimates of muscle strength. Some studies show the upper heritability estimates to be 96% (Huygens *et al.*, 2004*b*), whereas other studies predict heritability of strength to be as low as 14% (Beunen & Thomis, 2006). One reason for the lack of clarity in the hereditary estimates of muscle strength is that there are a lot of different measureable aspects of strength which are not consistent in the literature (Thomis *et al.*, 1998). Elbow extension and flexion has been shown to produce a higher heritability estimate than knee extension and flexion (Huygens *et al.*, 2004*b*, 2004*a*). Despite the wide range in heritability estimates of muscle strength there is a clear genetic component which explains the variance in the phenotype (Costa *et al.*, 2012). Unfortunately heritability studies do not actually inform about the causal genetic variant (Pitsiladis *et al.*, 2013). To allow the identification of the causal variant which underlie the heritability further studies, such as gene-association studies, must be conducted.

1.3 Genetic variation

The most commonly studied form of genetic variations studied are single nucleotide polymorphisms (SNPs). SNPs occur when one nucleotide is exchanged for another in the sequence of a gene. SNPs can occur in protein-coding exons, non-coding intronic regions of genes and intergenic regions (Barreiro *et al.*, 2008). The majority of SNPs fall within the introns of genes and thus typically do not have an effect on the function of the gene and are typically viewed as silent. There are occasions when intronic SNPs affect the splicing or expression of a gene and these SNPs may have an effect on the function of a gene (Choi *et al.*, 2008; Millar *et al.*, 2010). When a SNP falls within an exon it can be further classified as either non-synonymous or synonymous. Synonymous SNPs do not cause a change in the

amino acid and typically are silent or cause differential splicing of the gene. Non-synonymous SNPs lead to a change in amino acid and typically have the largest effects because they can either cause premature truncation of the protein or change the class of the amino acid potentially reducing the stability of the protein (Pollex & Hegele, 2007). For example, if a SNP causes a change from a hydrophobic amino acid to a hydrophilic amino acid, this can influence the stability of the protein depending on the cellular location (Strub *et al.*, 2004). The phenotypic consequence of a SNP will ultimately depend on the function of the gene that is affected. For example, a SNP variation in a gene that is functionally involved in the muscle contraction process may affect performance phenotypes.

The R577X polymorphism of the α-Actinin-3 (ACTN3) gene is one of the most studied and well documented which is associated with skeletal muscle phenotypes (for a detailed review see Berman & North, 2010). The X allele leads to a loss of function of ACTN3 that is associated with a reduced performance in power and sprint activities (Moran *et al.*, 2007; Walsh *et al.*, 2008). A study of elite athletes found a reduction of the X allele in sprint athletes and the complete absence of the X allele in Olympic athletes (Yang *et al.*, 2003). Several other SNPs have been associated with muscle size, strength and power phenotypes (for a detailed review see Ahmetov & Fedotovskaya (2012)). One of the key studies which has driven associations between genotypes and muscle phenotypes is the FAMuSS study (Thompson *et al.*, 2004). The FAMuSS study recruited a total of 1300 individuals and genotyped them for around 500 polymorphisms and found associations with several genetic loci and physical activity levels and muscle size, strength and body composition at baseline and in response to resistance training (Pescatello *et al.*, 2013).

Although SNPs are the most commonly studied class of genetic variant, they are not the only genetic variants that have been associated with functional consequences on phenotypes. The total heritability estimates of muscle performance traits have not been accounted for by SNP variation. Therefore, other classes of genetic variation, such as structural variants, may explain some of the missing heritability of muscle performance traits.

1.4 Structural Variants

Structural variation occurs when insertions, duplications or deletions of genetic material occur. The nature of a structural variant (SV) means it can not only impact several bases of a gene, but also several genes. Structural variants of the human genome can take the form of small deletions or insertions which impact up to 50 base pairs, through to large structural variants which can affect whole chromosomes (Boonpeng & Yusoff, 2013). The initial studies to detect genomic variation examined large structural variants because these studies were performed before the polymerase chain reaction (PCR) was invented. Without the technology to amplify the desired sequence of a gene, investigators relied on using microscope karyotyping and southern blot technology, which are limited to identifying large variants which affect large segments or whole chromosomes (Pollex & Hegele, 2007). The invention of PCR allowed the easy amplification of target gene sequences to identify SNPs, therefore the majority of studies moved into examining SNPs.

Several methods exist to identify SVs, each with associated advantages and disadvantages. These methods can be categorised into two main approaches, genome-wide scans for global identification of SVs and targeted scans of candidate genes. For genome-wide identification of SVs, the development of array based technology led to array-

Comparative Genomic Hybridization and SNP array based methods for detection of SVs. These array based methods use probes to compare the intensity of DNA samples to identify gains or losses of genetic material (Boonpeng & Yusoff, 2013). For the targeted scans approach, the traditional methods used are Southern blotting and fluorescence in situ hybridization (FISH) (Pollex & Hegele, 2007). Both southern blots and FISH are infrequently used because of the drawbacks associated with each of the methods. Southern blots require a high quantity of DNA and is a laborious process, while FISH is not suited to identifying a partial deletion of genes (Hoebeeck *et al.*, 2007), because of resolution limitations (Boonpeng & Yusoff, 2013). The most common and simplest method for SV detection is to use quantitative PCR (qPCR) to compare the fluorescence of a target sequence compared to a reference assay (Boonpeng & Yusoff, 2013). One issue with qPCR based SV detection is the use of a single gene as the reference assay (e.g. RNASE P or TERT) because SVs could affect the assay which can lead to the incorrect determination of copy number of the target gene (Table 1).

Table 1 - Problems that can develop through the use of a single gene for the relative quantification of structural variants.

Reference Assay Copy Number	Target Gene Copy Number	Calculated Copy Number of Target	Correct copy Number for Target Gene
1	2	4	No
2	2	2	Yes
3	2	1.33	No

1.5 Copy number variation

Generally it is accepted that individuals have two copies of every gene, one inherited from their mother and one inherited from their father; however, it has been discovered that

the variation in copy number of a gene because of duplications and deletions of genetic material is widespread (Yang *et al.*, 2013). Advances in genotyping and the human genome project allowed the investigation of gene copy numbers at a subchromosomal level (Lobo, 2008). A SV of DNA of 1000 nucleotides or more which differs in copy number from a reference genome is classified as a copy number variant (CNV) (Redon *et al.*, 2006). In 2004, the first papers to identify the widespread extent of CNVs throughout the human genome were published (Sebat *et al.*, 2004; Iafrate *et al.*, 2004). Both of these studies identified approximately 250 regions of copy number losses or gains in two different cohorts of unrelated individuals. A larger study of 270 individuals identified 1447 CNV regions throughout the human genome which affected 360 mega bases of DNA; equivalent to 12% of the total human genome and between 6-19% of any chromosome (Redon *et al.*, 2006). A more recent study identified that on average 5% of the genome is affected by CNVs (Conrad *et al.*, 2010). These CNVs can either be inherited from parents or rise from de novo mutations, where both parents have a normal copy number at the region of DNA in question. Studies have shown that monozygotic twins can differ in copy number for genes (Hastings *et al.*, 2009) and the same individuals can have a different copy number depending on the tissue examined (Piotrowski *et al.*, 2008).

With the vast distribution of CNVs throughout the genome in a range of different organisms it is expected that they will contribute to the phenotypic differences between individuals. The main way CNVs will influence a phenotype will be through an alteration of gene dosage, causing a disturbance in coding sequences or disturbing the long-range regulation of genes through alteration of chromatin environment (Stranger *et al.*, 2009). Between 85-95% of CNVs examined have been associated with changes in the expression of the affected genes (Stranger *et al.*, 2009; Henrichsen *et al.*, 2009). Deviation in gene copy

number associates with changes in the amount of the final protein levels. The proteins which do not show altered product tend to be components of complexes (Tang & Amon, 2013). On dosage sensitive genes, which have no compensatory mechanism, the presence of a duplication or deletion will directly increase or decrease the expression of the gene, respectively (Pollex & Hegele, 2007). When gene duplications occur there are several consequences for gene expression depending on the inclusion or absence of regulatory elements. If complete duplication of a gene including the regulatory region occurs there will be a greater increase in expression than if partial duplication of a gene occurs. Duplication of a gene can reduce expression level if the duplicated region ends up in different regulatory regions, such as near insulators or repressors, a reduction in expression will be observed. Duplications also can reduce expression if they prevents access to the regulatory machinery preventing efficient transcription (Henrichsen *et al.*, 2009).

The association of CNVs with changes in phenotypes is not a recent development. One of the earliest functional CNVs to be identified occurred 78 years ago when a duplication in the Bar gene in Drosophila was discovered to narrow the eye field (Bridges, 1936). The change in copy number in any number of genes is not well tolerated and often leads to the development of disease (Hastings *et al.*, 2009). Most studies examining CNVs in humans have examined the deleterious nature of CNV on phenotypes in diseased populations and have led to various associations between CNVs and disease phenotypes (Almal & Padh, 2012). Various neurological disorders including Parkinson's disease (Singleton *et al.*, 2003; Pankratz *et al.*, 2011), Alzheimer's disease (Sleegers *et al.*, 2006) and Autism (Weiss *et al.*, 2008) have been associated with CNVs. Numerous forms of cancers, such as prostate, breast and colorectal have also been associated with CNVs (Yang *et al.*, 2013). Several chronic health related diseases have also been associated with CNVs including various cardiovascular

diseases (Pollex & Hegele, 2007) and metabolic disorders including both type I (Grayson *et al.*, 2010) and type II (Jeon *et al.*, 2010) Diabetes. There have been many CNVs found on known candidate genes for body mass. For example, the deletion of a 2.1Mb region which spans the UPC1 and IL15 genes increases the risk of obesity (Wang *et al.*, 2010). CNV has previously been associated with lean body mass. Individuals with two copies of CNV2073 had 6.9% greater lean body mass than individuals with 3 copies and 11.2% more than individuals with 4 copies. One gene which is located in the region of CNV2073 is Gremlin1 which has already been reported as a candidate for lean body mass by linkage studies and molecular function studies (Hai *et al.*, 2012). Gains of two CNVs (CNV1191 and CNV2580) have also associated with decrease in appendicular lean body mass (Ran *et al.*, 2014).

Despite most of the associations with CNVs being with diseased phenotypes, CNVs have been shown to occur with healthy populations and have advantageous consequences. Salivary amylase (AMY)1 is the most common example of a gene which has CNV with beneficial effects. In populations with high starch diets it has been shown that there is an increase in the AMY1 copy number and elevated expression of the salivary amylase protein which is responsible for starch hydrolysis (Perry *et al.*, 2008). Increased copy number for AMY1 also has been associated with a significantly reduced BMI and risk of obesity (Falchi *et al.*, 2014). CNVs of other genes have been associated with beneficial phenotypes such as delayed progression of spinal muscular atrophy with an increase in the number of copies of the survival motor neuron (SMN)2 gene (Elsheikh *et al.*, 2009). Patients with three or four copies of SMN2 have been shown to produce higher force for grip, elbow flexion, and knee extension and flexion compared to the patients with two copies; none of the increases reached statistical significance (Febrer *et al.*, 2010). These studies show that CNVs can consequences in clinical populations and in the general population. To date no studies have investigated the

impact of CNVs with muscle performance in athletes. We would predict that because of the previous associations with disease, that CNVs in genes related to athletic performance will have phenotypic consequences for muscle function.

1.6 Myosin Light Chain Kinase

The Myosin Light Chain Kinase (MYLK) genes encode for the MYLK proteins which are responsible for the phosphorylation of the regulatory light chain of myosin (RLC). Western blot analysis has shown that the different MYLK proteins differ in the site of expression (Zhi *et al.*, 2005). MYLK1 has been shown to be expressed ubiquitously throughout all tissues, however the main site of expression of the kinase is in smooth muscle and therefore it has been termed the smooth muscle MYLK (smMLCK). MYLK2 has been shown to express the skeletal muscle MYLK (skMLCK) specifically in skeletal muscle and predominantly in the type II fast twitch fibres (Zhi *et al.*, 2005; Herring *et al.*, 2010; Kamm & Stull, 2011). MYLK3 encodes for the cardiac MYLK (cmMLCK) protein which is expressed mainly in the cardiac myocytes (Takashima, 2009). The phosphorylation of the RLC of myosin is important in the contraction process for all three of the muscle types (smooth, skeletal and cardiac) (Takashima, 2009). In general MYLK controls the phosphorylation of the RLC of myosin to allow interaction between myosin and actin filaments in the contraction of muscle fibres (Stull *et al.*, 2011).

Each of the MYLK proteins have different roles specific to the type of muscle it is associated with. In smooth muscle, an increase in intracellular calcium leads to the activation of smMLCK by calmodulin (Kamm & Stull, 2001). The phosphorylation of the RLC by smMLCK has been shown to be sufficient to initiate the contraction process in smooth

muscles through the activation of myosin ATPase (Figure 1)(Kamm & Stull, 2001; Murthy, 2006). In both skeletal and cardiac muscle MYLK activation is not the key step in the initiation of the contraction process because it is not the calcium sensor, instead that role is played by troponin. The binding of calcium to troponin causes a conformational change, allowing interaction between the myosin cross-bridges and actin filaments which leads to the development of force (Farah & Reinach, 1995). In smooth muscle, RLC phosphorylation by smMLCK affects the myosin ATPase activity, whereas in skeletal and cardiac muscle there is no impact on ATPase activity by the specific MLCK. In skeletal muscle, the phosphorylation of RLC by skMLCK leads to changes in properties of the myosin cross-bridges (Figure 1)(Kamm & Stull, 2011). In mice with ablated MYLK2, it has been shown that there is a role for RLC phosphorylation in the potentiation of isometric twitch force. The knockout mouse displays a reduction in the phosphorylation of the RLC and produces significantly lower twitch force (Zhi *et al.*, 2005; Gittings *et al.*, 2011). There is also evidence of some RLC phosphorylation in the knockout mouse which suggests there could be some redundancy between skMLCK and smMLCK (Takashima, 2009). The functional role of MYLK1 and MYLK2 in the contraction of muscle and the suggestion of a potential functional redundancy for skMLCK by smMLCK makes both of these genes candidate genes for genetic variation which is functionally involved in muscle strength and power phenotypes.

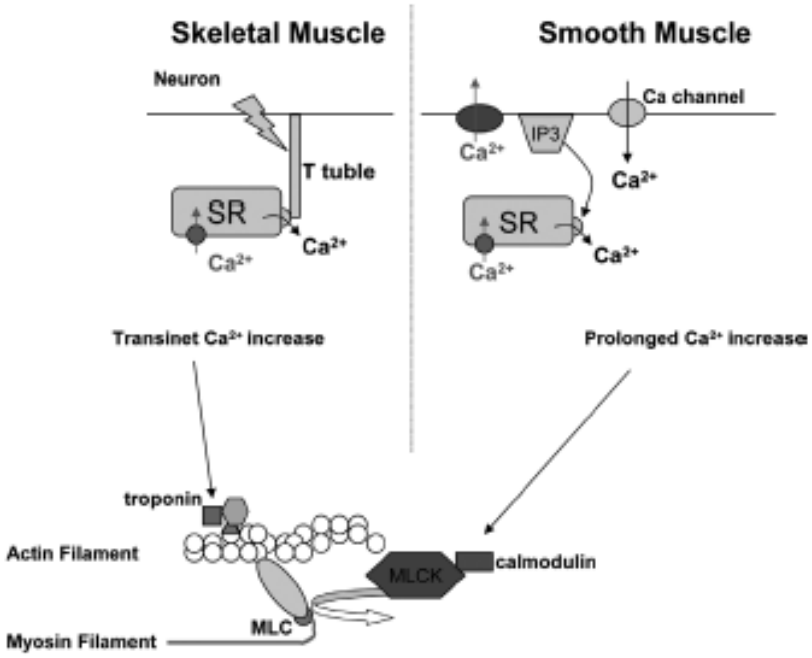

Figure 1 - Role of MLCK in the contraction of skeletal and smooth muscle (adapted from Takashima (2009)). SR; sarcoplasmic reticulum, Ca^{2+}; calcium ions, MLCK; myosin light chain kinase, sk; skeletal muscle, sm; smooth muscle, MLC; myosin regulatory light chain.

Previous SNP association studies attempted to examine if variation in MYLK2 is associated with muscle function. However, the SNPs selected in these studies are found to encode for MYLK1 (C49T (rs2700352) and C37885A (rs28497577)) (Clarkson *et al.*, 2005; Deuster *et al.*, 2013). Despite both of the studies attempting to examine variation in MYLK2 but instead examining MYLK1, associations with strength phenotypes were found. The first study to examine SNP variation in MYLK1 and muscle phenotypes studied 157 patients who performed a muscle damage protocol of the elbow flexors to investigate the genetic determinants of exertional muscle damage (Clarkson *et al.*, 2005). The two muscle strength phenotypes examined were baseline strength and greater strength loss following eccentric exercise of the elbow flexors. A key finding of the study was contrary to the hypothesis as they found patients homozygous for the T allele of C49T produced significantly higher baseline strength values. They also found that patients homozygous of the C allele of

C37885A showed significantly lower post exercise strength loss. A second study examined the risk of developing exertional rhabdomyolysis in 47 disease sufferers and 134 control individuals to identify polymorphisms associated with the disease (Deuster *et al.*, 2013). Two of the polymorphisms that were investigated were the same MYLK1 SNPs as the Clarkson study. They did not find any association with C49T which is consistent with the previous study only finding a relationship with baseline strength. Individuals homozygous for the A allele of MYLK1 C37885A were determined to be 5 times more likely to suffer from exertional rhabdomyolysis than individuals who were CC homozygotes (Deuster *et al.*, 2013), which is consistent with the findings that the C allele causes lower strength loss following exercise (Clarkson *et al.*, 2005). The results of the two studies show that MYLK1 C49T and C37885A are both associated with muscle function changes in clinical populations (Clarkson *et al.*, 2005; Deuster *et al.*, 2013); however these SNPs have yet to be examined in an athletic cohort.

Unpublished data from our laboratory examined the same SNPs of MYLK1 as previous studies (Clarkson *et al.*, 2005; Deuster *et al.*, 2013) in a cohort of athletes and control individuals in an attempt to find if the associations with muscle performance phenotypes in clinical populations remain when examined in an athletic population. Three SNPs of MYLK2 were also examined to test if variation in the skeletal muscle specific gene also associated with muscle performance phenotypes. No muscle strength phenotypes were associated with the SNPs of MYLK2; however, associations were made with isokinetic test performance and the MYLK1 SNPs. Given that SNP variation has been associated with functional changes in muscle performance, we hypothesise that CNVs will affect the function of MYLK1 and therefore be associated with muscle strength phenotypes.

The antisense gene of MYLK1, called MYLK-antisense 1 (MYLK-AS1) is potentially another candidate gene for functional variation associated with athletic performance. Antisense genes are thought to functionally act to regulate gene expression, however the specific role of these mRNA genes is not fully understood (Werner, 2013). These antisense transcripts form duplex RNA through hybridization with the sense transcript. These interactions can reduce the gene expression and decrease the levels of protein synthesis by preventing the translational machinery access to the sense transcript (Kumar & Carmichael, 1998). Another potential role of antisense genes is to control the splicing of the associated sense gene, by preventing access to splice sites (Morrissy *et al.*, 2011). Some antisense genes have been shown to stabilise the expression of the sense genes through competition for microRNA binding sites and RNA motifs (Werner, 2013). Several antisense genes have been shown to be functional however the mechanisms underlying the function have yet to be identified. It is thought that MYLK-AS1 may act to reduce the expression of MYLK1 leading to a reduction in RLC phosphorylation and decreased muscle performance.

The Database of Genomic Variants (MacDonald *et al.*, 2014) identified the presence of CNVs which span the whole of the MYLK1 and MYLK-AS1 genes. There have been no CNVs detected which span the whole of the MYLK2 gene. The studies which identified the presence of CNVs on MYLK1 and MYLK-AS1 did not include any functional measures therefore the phenotypic consequences of changes in copy number in these genes is unknown (Wong *et al.*, 2007; Itsara *et al.*, 2009).

1.7 Aims & Objectives

The aim of the present study was to assess copy number variation in a cohort of athletes and control individuals for the first time and examine if it is associated with strength. MYLK1 and MYLK-AS1 have been selected as the target genes to examine if some of the missing heritability of muscle strength and power traits can be explained by CNVs in genes previously associated with functional SNP variation. There are three main objectives of the project: (1) to examine the extent of variation in MYLK1 and MYLK-AS1 copy number in athletes and control; (2) to examine if there is a different relationship between MYLK1 and MYLK-AS1 copy number between different groups of athletes (endurance and strength/power athletes) and controls; (3) to investigate how the copy number of the two target genes associates with quantitative phenotypes associated with strength and power. We hypothesise that increases in MYLK1 copy number will be overrepresented in the strength athletes and consequently also be associated with increased performance in tests of strength and power phenotypes. We expect to find the opposite in MYLK-AS1 with an underrepresentation in the strength athletes and decreased performance in measures of strength and power. The result of this study will further the understanding of genetic variation which influences muscle function.

Chapter 2 - Materials and Methods

2.1 Ethical Approval

Ethical approval was obtained for the genetics of Lithuanian athletes (GELA) study by the Lithuanian State Bioethics Committee and the University of Stirling, School of Sport Research Ethics Committee. The study abides by the principles outlined in the Declaration of Helsinki. Informed consent in writing was obtained from each participant prior to participation in the study.

2.2 Participant characteristics

The present study utilised the pre-established GELA cohort which consists of native Lithuanians (n=447) from the Lithuanian Sports University (Kaunas, Lithuania). All participants were males (17-37 years old) and where related individuals were recruited only the individual who competed at the highest level was retained for analysis. All participants underwent phenotype determination for a range of standard physiological tests (see: phenotype data). To confirm all participants were performing close to their peak at the time of phenotype determination, any participants unable to perform the physiological tests due to injury or disease were excluded from the study. Any athletes who could not be categorised as strength-power (STP), endurance (END) or they no longer performed at their top level were also excluded from the study. A total of 407 participants remained for further analysis following the exclusions of the participants who did not meet the necessary criteria (n=40).

The participants were divided into 3 separate groups: control individuals (n=197), strength and power athletes (STP; n=126) and endurance athletes (END; n=84). Descriptive characteristics of all participants can be found in Table 2. All athletes trained a minimum of twice a week, whereas the control participants did not take part in organised physical activity more than twice a week and did not compete in competitive sports unless it was a low impact sport. STP included weightlifters, combat athletes, field athletes (discus, javelin and long jump) gymnasts, team sport individuals (handball, volleyball, basketball and football), short distance swimmers and runners, a motocross-rally driver and a high-diver. END included long-distance runners, cyclists, triathletes, a rower, modern pentathletes, orienteers, skiers and walkers. Each of the athlete subgroups was further divided by highest level of competition. International athletes (n=14) were the participants who represented Lithuania at the Olympic Games or World Championships. National athletes (n=97) were ranked in the top 3 for their sport or have been included in a Lithuanian national sport squad but did not compete internationally. Regional athletes (n=63) were defined as athletes that were members of their local area squad. Other athletes (n=36) were a mixture of athletes who did not achieve any of the above levels or chose not to compete but trained regularly. The number of athletes who performed each sport and the level of performance of these athletes can be found in Appendix A.

Table 2 - Descriptive characteristics of the groups of participants. Values are means ± 95% CI. FFM; fat free mass, STP; strength and power athletes, END; endurance athletes

Phenotype	Control (n=197)	Athletes (n=210)	STP (n=126)	END (n=84)
Age (yrs)	23.98 ± 0.60	21.77 ± 0.46	21.83 ± 0.57	21.68 ± 0.78
Height (cm)	180.29 ± 0.83	181.65 ± 2.58	183.27 ± 4.23	179.27 ± 1.17
Body mass (kg)	77.00 ± 1.50	76.90 ± 1.62	81.30 ± 2.26	70.35 ± 1.44
Body Fat (%)	15.80 ± 0.73	12.99 ± 0.48	14.12 ± 0.65	11.29 ± 0.51
FFM (kg)	64.28 ± 0.95	66.34 ± 1.63	69.07 ± 2.55	62.37 ± 1.20
Training experience (yrs)	N/A	8.04 ± 0.80	8.83 ± 1.05	6.87 ± 1.17
Training volume (hr/wk)	N/A	11.4 ± 1.13	10.72 ± 1.55	12.36 ± 1.6

2.3Phenotype data

Participants were required to attend the laboratory for testing on three separate occasions having rested for at least 12 hours and fasted for at least 2 hours. On entry to the laboratory on the first day all participants were assigned a unique ID code to allow all data to be stored in an anonymous manner. All participants had a 10 ml venous blood sample extracted into an EDTA container and frozen at -80 ^{o}C prior to DNA extraction. During the laboratory visits the participants were required to fill out a number of questionnaires and complete a range of standardised physiological tests, which were carried out by trained individuals. The laboratory visits allowed detailed phenotypes to be collected including measures of anthropometrics, body composition, endurance performance, strength performance and cardiac size and function. For the current analysis a subset of the phenotypes which are indicative of strength and power performance. The selected tests included measures of isolated muscle group force production such as hand grip strength and isokinetic dynamometry for extension and flexion of the right and left legs and arms at a

range of speeds (30-180 °/s) and performance in whole body movements such as number of pull-ups, countermovement jump and static jump height and time to complete a 30 metre sprint (Details on the tests performed in Appendix B).

2.4 DNA Extraction

All extractions were performed at the University of Stirling. Genomic DNA was extracted from 200 µl of whole blood using silica columns (Macherey-Nagel GmbH & Co. KG, Germany) and extraction buffers (Qiagen Ltd, UK) using the QIAamp Spin DNA extraction protocol. DNA yield and quality was checked by absorbance on a Synergy HT microplate reader (BioTek, USA) with Gen5.0 software. Samples with a yield of less than 10 ng/µl were re-extracted (concentration of DNA samples can be found in Appendix C). All DNA samples were stored at -20 ^{o}C prior to analysis.

2.5 Copy Number Assays

To analyse the copy number of MYLK1 and MYLK-AS1, a multiplexed quantitative real-time polymerase chain reaction (qPCR) method was used. Primer and hydrolysis probe assays (Integrated DNA Technologies Inc, USA) were designed to amplify a region which lies in the introns of the genes of interest. Care was taken to ensure that no SNPs had been reported previously in European populations within the primer and probe sequences. The Basic Local Alignment Search Tool (BLAST, National Centre for Biotechnology Information) was used to ensure that the primers and probes were specific to the genes of interest and did not recognise any other sequences. The MYLK1 assay contains the forward primer (5'-ACATACACAAGGTCAGTCACG-3'), reverse primer (5'-

GAGTCTGAACTCACAACGGTAG -3') and Cy5 labelled probe (5'-CAAGCCACTGATGAAGTGATGGCA-3') which amplifies a 99 base pair region. The MYLK-AS1 assay contains the forward primer (5'-TCCAGAGAGAGAGAAACCTTAGTT-3'), reverse primer (5'- GAGATGGTCCAGGATAGTAAAGATAAAG-3') and 6-FAM labelled probe (5'-TGTCCAGGAAGGGAGGAAGAAGAGT-3') which amplifies a 101 base pair region. Figure 2 displays the location of these two amplicons. For determination of gene copy numbers the Type-it CNV probe PCR kit + qC kit was used including the multicopy Type-it CNV Reference Probe Assay Solution (Qiagen Ltd, UK) of a known copy number labelled with a MAX fluorescence probe. The use of this reference assay was selected as the control assay because it has a copy number of 36 which makes it more robust to CNVs than traditional assays which use single copy genes (Table1). The fluorescence emission spectra of the 6-FAM (494-518), MAX (524-557) and Cy5 (643-667) dyes are all close together, therefore, colour compensation was performed to optimise fluorescence detection to prevent any crossover fluorescence from the other probes being picked up.

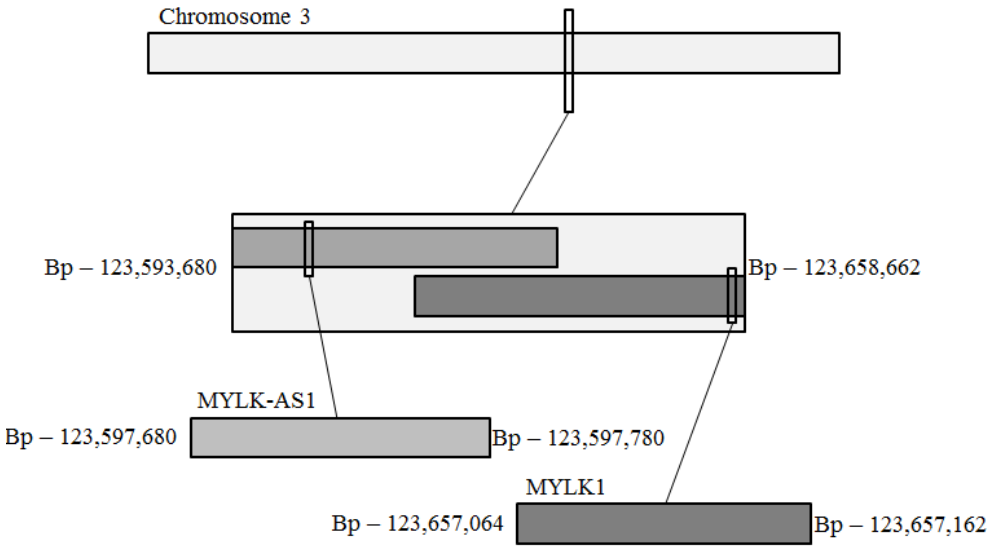

Figure 2 – Location of the MYLK1 and MYLK-AS1 sequences on Chromosome 3 according to the Genome Reference Consortium Human Build 38. Bp; base pair location on chromosome 3, MYLK1; myosin light chain kinase1, MYLK-AS1; myosin light chain kinase-antisense1.

All samples were run in triplicate and each plate contained a non-template control also run in triplicate. A reaction volume of 12.5 µl was used for all reactions, containing 6.25 µl of 2x Type-it CNV Probe PCR Master Mix (Qiagen Ltd, UK), 0.5 µl of 25x Ref assay, 0.5 µl of 25x MYLK1 assay, 0.5 µl of 25x MYLK-AS1 assay, 2.75 µl of nuclease-free water and 2 µl of DNA normalised to 5 ng/µl. All reactions were performed on the same LightCycler 480 (Roche Diagnostics Ltd, UK) system and included activation at 95 ^{o}C for 5 min followed by 40 cycles of alternative 30 s periods at 95 ^{o}C and 60 ^{o}C as recommended by the Type-it CNV probe protocol.

2.6 Copy number determination

The median absolute deviation method was used to identify any outlier crossing point (Cp) within each triplicate for each assay. The mean of the remaining Cps in each triplicate

was calculated to give the corrected Cp for each sample in the cohort. The median absolute deviation was also used to identify any samples from the cohort which needed to be repeated because of outlying corrected values. The difference between the Cp of genes of interest and the reference assay was calculated to allow determination of the copy number for each participant.

In qPCR reactions, primers which amplify at an efficiency of 90-110% are acknowledged as acceptable for the amplification of targets (D'haene *et al.*, 2010). There are a number of statistical methods which can be used to determine qPCR efficiency all of which can produce a different result because of the different algorithms used (Ruijter *et al.*, 2013). Small changes in reaction efficiency can dramatically alter copy number estimates (Yuan *et al.*, 2007). Standard curves of serial dilutions and LinRegPCR (Ramakers *et al.*, 2003) were used to calculate the efficiencies of the MYLK1 and MYLK-AS1 assays. Different values (range 85-107%) were obtained for both of these measures which would change the number of copies of these genes. The normal copy number for most genes is two and in previous studies only four out of ~1700 individuals have been detected with alterations (duplications or deletions) of copy number (Wong *et al.*, 2007; Itsara *et al.*, 2009). In the present cohort, we calculated that efficiencies of 105.7% and 99.8% for MYLK1 and MYLK-AS1 respectively would lead to a median copy number of two. These efficiencies are within the range that we calculated; therefore we scaled the obtained Cp values using these efficiencies. It should be noted that the scaling of these data will not affect the associations with quantitative phenotypes. All statistical analyses were performed with, and without, the scaling of data and similar results were obtained. To generate the copy number for each participant, copy number was grouped into the nearest integer as has been previously described (Gonzalez *et al.*, 2005; Cantsilieris *et al.*, 2012). The grouping of copy number may reduce the power to detect

associations, however it leads to clearer interpretation of associations and similar results were obtained when correlations were performed on the raw copy number or integer copy number.

2.7 Statistical Analysis

All statistical analysis was carried out using Minitab (version 16, Minitab, State College, PA). All phenotype data were tested for normality using the Ryan-Joiner test (Ryan, 1974). Any data that were not normal were transformed using the Box-Cox transformation to give better approximation of the normal distribution. Data were back-transformed for display in figures and tables, thus all values are means ± 95% confidence intervals unless otherwise stated.

2.7.1 Comparison of copy number / genotype between groups

Chi-squared contingency tables (corrected using the Yates correction for continuity (Yates, 1934) where appropriate) were used to determine if the athletic groups differed from each other or from the control group in copy number for each of the target genes. Differences between the groups were examined using odds ratios (OR) to identify the likelihood of an individual being assigned to a particular athlete group or performance level achieved based on copy number. The significance of the OR was determined using chi squared contingency tables and 95% confidence intervals (CI). Significance was accepted as chi-squared $p<0.05$.

2.7.2Association of copy number with physiological variables

All phenotype data were z-scored using the overall mean and standard deviation. Phenotype data were separately z-scored within each of the subgroups to account for group differences. Pearson's correlation analysis was performed using the z-scored phenotype data with and without group corrections to identify associations between copy number and physiological variables. The Benjamini-Hochberg (BH) correction for multiple testing was applied to control for false discovery rate separately for isolated muscle and whole body performance tests (Benjamini & Hochberg, 1995). Pearson's correlations were accepted as significant where p-value < α-value determined in BH correction. For the significant correlations, two tailed T-tests were used to compare the means in quantitative phenotypes for each copy number. T-test significance was accepted as $p<0.05$.

Chapter 3 - Results

3.1 Copy number detection in each group

Copy number determination was achieved for 405 of the individuals (99.5% success). The same two individuals failed for both target genes and the reference assay. Out of the 405 individuals examined, 12 (3%) individuals have variations of MYLK1 away from the normal copy number of two, whereas 18 (4.5%) individuals have variations of copy number for MYLK1-AS1. Only one control had a deviation of copy number for MYLK1 which is at the frequency expected (Wong *et al.*, 2007; Itsara *et al.*, 2009); however, controls with deviations in copy number for MYLK-AS1 are found at a greater frequency than anticipated. It should be noted, however, that previous studies used to determine expected frequencies used a range of ethnicities which could differ in allele frequencies (Wong *et al.*, 2007; Itsara *et al.*, 2009). In the group of athletes (N=208), 11 and 12 have alterations away from the normal copy number of two for MYLK1 and MYLK-AS1 respectively. Deviations away from a copy number of two have been shown to be a more regular occurrence in STP than END for both of the target genes. The number of deletions and duplications within each of the groups can be found in Table 3.

Table 3 – The copy number of the individuals within each of the groups. Values are the total number with the percentage of the group in brackets. CN; copy number, STP; strength and power athlete, END; endurance athlete

	MYLK1 CN (%)			MYLK-AS1 CN (%)		
	1	2	3	1	2	3
Total (n=405)	3 (0.74)	393 (97.04)	9 (2.22)	3 (0.74)	387 (95.56)	15 (3.70)
Athlete (n=208)	3 (1.44)	197 (94.71)	8 (3.85)	2 (0.96)	196 (94.23)	7 (3.41)
STP (n=124)	2 (1.61)	116 (93.55)	6 (4.84)	1 (0.81)	118 (95.16)	5 (4.03)
END (n=84)	1 (1.19)	81 (96.43)	2 (2.38)	1 (1.19)	81 (96.43)	2 (2.38)
Control (n=197)	0 (0.00)	196 (99.49)	1 (0.51)	1 (0.51)	188 (95.43)	8 (4.06)

3.2 Copy number differences between groups

A trend exists where the copy number of MYLK1 in the athletes differs to the controls (χ^2 =5.150, p=0.076). When athlete subgroups are examined, the distribution of copy numbers is significantly different between STP and the controls (χ^2 =6.103, p=0.047). END did not differ from the controls (χ^2 = 0.828, p=0.661) or STP (χ^2 =0.436, p=0.804) for MYLK1 copy numbers. OR display that athletes are less likely to have two copies (p=0.005) of MYLK1 and there is a trend where athletes are more likely to have three copies (p=0.051) compared to the controls. STP are less likely to have two copies (p=0.003) and more likely to have three copies (p=0.028) than controls (Table 4). There were no associations between MYLK-AS1 copy number and any of the groups (χ^2 = 0.114-0.268; p=0.874-0.945).

Table 4 - Odds Ratios for athlete group vs copy number for MYLK1. Odds Ratio of N/A is because none of the controls have a copy number of one for MYLK1. STP; strength and power athletes, CI; Confidence interval

Comparison	Copy number	p-value	Odds Ratio (95% CI)
Athletes vs Controls	1	0.264	N/A
	2	**0.005**	**0.09 (0.01-0.71)**
	3	0.051	7.84 (0.97-63.27)
STP vs Controls	1	0.287	N/A
	2	**0.003**	**0.07 (0.01-0.60)**
	3	**0.028**	**9.97 (1.19-83.81)**

In the next stage of analysis, the highest level of competition was investigated for associations between copy numbers of the target genes. For MYLK1, the only association found with level of performance was between Regional level athletes and the controls (χ^2 =6.618; p=0.037). The Regional athletes are less likely to have two copies (p=0.003 [OR 0.06; 95% CI:0.01-0.5]) and more likely to have three copies (p=0.013 [OR 13.75; 95% CI:1.51-125.52]) than the controls. No associations were found for highest level of competition and MYLK-AS1 copy number (χ^2 =0.055-1.644, p=0.440-0.973).

3.3 Copy number associations with quantitative measurements

To account for the difference in training of the participants, the relationship between copy number and quantitative phenotypic measurements has been investigated with and without corrections for subgroup (Table 5 and 6). To test for associations which could potentially confound other analyses, correlations were made between MYLK1 / MYLK-AS1 copy number and age, height, body mass, training history and volume. No significant

correlations existed between any of these variables and MYLK1 / MYLK-AS1 copy number before or after correction for subgroup (p=0.091-0.878).

Copy number of the target genes was investigated for associations with selected phenotypes associated with strength and power performance for isolated muscle groups and whole body movements. To assess the correlation between these phenotypes and copy number, Pearson's correlations were performed with BH correction for multiple testing was applied to reduce the risk of type 1 statistical errors. Overall in the whole cohort MYLK1 copy number was associated with strength and power phenotypes, in isolated muscle groups and whole body movements (Table 5). The average variance explained by these significant correlations is 1.5% with the highest variance being explained by isokinetic extension of the elbow (2.76%). After correction for subgroup the associations for counter movement jump and sprint performance are lost whilst the remaining associations are weakened (Table 5). The average variance explained by the group corrected correlations is 1.35 %. The highest percentage of the variance is still found with the isokinetic extension of the elbow, however the variance drops down to 2.34 %. None of the strength and power phenotypes correlated with copy number for MYLK-AS1 with or without the correction for subgroup (Table 6).

Table 5 - Pearson's correlation coefficient results for MYLK1 copy number and quantitative variables before and after correction for subgroup. Significant (p<0.05) values are highlighted in bold. Values marked by * have a p-value lower than the acceptable alpha value according to the BH correction. kgFFM; kilogram of fat free mass, IKR, isokinetic right; IKL, isokinetic left

Trait	MYLK1					
	Before Subgroup Correction			After Subgroup Correction		
	p-value	**Coefficient**	**R^2 (%)**	**p-value**	**Coefficient**	**R^2 (%)**
Number of pull-ups	0.08	0.089	0.792	0.725	0.018	0.032
Counter movement jump height (cm)	**0.049**	**0.099**	**0.980**	0.132	0.076	0.578
Static jump height (cm)	**0.009***	**0.132**	**1.742**	**0.022***	**0.116**	**1.346**
Sprint 0-10 (s/kgFFM)	**0.022***	**-0.116**	**1.346**	0.077	-0.09	0.810
Sprint 10-20 (s/kgFFM)	**0.018***	**-0.121**	**1.464**	0.055	-0.097	0.941
Sprint 20-30 (s/kgFFM)	**0.028**	**-0.112**	**1.254**	0.076	-0.09	0.810
Sprint Total (s/kgFFM)	**0.020***	**-0.118**	**1.392**	0.064	-0.094	0.884
Total Handgrip (kg)	**0.029**	**0.111**	**1.232**	**0.046**	**0.102**	**1.040**
IKR knee extension torque (30 °/s)	0.081	0.091	0.828	0.150	0.075	0.563
IKL knee extension torque (30 °/s)	**0.026**	**0.116**	**1.346**	**0.046**	**0.104**	**1.082**
IKR knee extension torque (90 °/s)	**0.024**	**0.117**	**1.369**	0.071	0.094	0.884
IKL knee extension torque (90 °/s)	**0.007***	**0.139**	**1.932**	**0.024**	**0.117**	**1.369**
IKR knee extension torque (180 °/s)	**0.045**	**0.104**	**1.082**	0.128	0.079	0.624
IKL knee extension torque (180 °/s)	**0.020**	**0.121**	**1.464**	0.064	0.096	0.922
IKR knee flexion torque (30 °/s)	0.220	0.064	0.410	0.399	0.044	0.194
IKL knee flexion torque (30 °/s)	0.199	0.067	0.449	0.410	0.043	0.185
IKR knee flexion torque (90 °/s)	0.137	0.077	0.593	0.303	0.054	0.292
IKL knee flexion torque (90 °/s)	0.156	0.074	0.548	0.352	0.048	0.230
IKR knee flexion torque (180 °/s)	0.064	0.096	0.922	0.185	0.069	0.476
IKL knee flexion torque (180 °/s)	0.084	0.090	0.810	0.226	0.063	0.397
IKR elbow extension torque (30 °/s)	**0.009***	**0.136**	**1.850**	**0.021**	**0.120**	**1.440**
IKL elbow extension torque(30 °/s)	**0.009***	**0.135**	**1.823**	**0.024**	**0.117**	**1.369**
IKR elbow extension torque (90 °/s)	0.177	0.071	0.504	0.225	0.063	0.397
IKL elbow extension torque (90 °/s)	**0.032**	**0.111**	**1.232**	**0.046**	**0.104**	**1.082**
IKR elbow extension torque(180 °/s)	**0.001***	**0.166**	**2.756**	**0.003***	**0.153**	**2.341**
IKL elbow extension torque (180 °/s)	**0.028**	**0.114**	**1.300**	**0.04**	**0.107**	**1.145**
IKR elbow flexion torque (30 °/s)	**0.014***	**0.128**	**1.638**	**0.033**	**0.111**	**1.232**
IKL elbow flexion torque (30 °/s)	**0.026**	**0.115**	**1.323**	0.063	0.096	0.922
IKR elbow flexion torque (90 °/s)	0.292	0.055	0.303	0.451	0.039	0.152
IKL elbow flexion torque (90 °/s)	0.121	0.080	0.640	0.199	0.067	0.449
IKR elbow flexion torque (180 °/s)	0.456	0.039	0.152	0.536	0.032	0.102
IKL elbow flexion torque (180 °/s)	**0.009***	**0.135**	**1.823**	**0.019**	**0.121**	**1.464**

Table 6 - Pearson's correlation coefficient results for MYLK1-AS1 copy number and quantitative variables before and after correction for subgroup. Significant (p<0.05) values are highlighted in bold. Values marked by * have a p-value lower than the acceptable alpha value according to the BH correction. kgFFM; kilogram of fat free mass IKR, isokinetic right; IKL, isokinetic left

Trait	MYLK-AS1					
	Without Group Correction			With Group Correction		
	p-value	**Coefficient**	**R^2 (%)**	**p-value**	**Coefficient**	**R^2 (%)**
Number of pull-ups	0.807	0.012	0.014	0.133	0.077	0.593
Counter movement jump height (cm)	0.795	0.013	0.017	0.869	0.008	0.006
Static jump height (cm)	0.583	0.028	0.078	0.575	0.028	0.078
Sprint 0-10 (s/kgFFM)	0.913	-0.006	0.004	0.918	0.005	0.003
Sprint 10-20 (s/kgFFM)	0.849	-0.010	0.010	0.934	-0.004	0.002
Sprint 20-30 (s/kgFFM)	0.731	-0.017	0.029	0.781	-0.014	0.020
Sprint Total (s/kgFFM)	0.843	-0.010	0.010	0.952	-0.003	0.001
Total Handgrip (kg)	0.272	0.056	0.314	0.412	0.042	0.176
IKR knee extension torque (30 °/s)	0.911	0.006	0.004	0.953	-0.003	0.001
IKL knee extension torque (30 °/s)	0.094	0.087	0.757	0.100	0.086	0.740
IKR knee extension torque (90 °/s)	0.768	0.015	0.023	0.924	0.005	0.003
IKL knee extension torque (90 °/s)	0.949	0.003	0.001	0.849	-0.010	0.010
IKR knee extension torque (180 °/s)	0.699	0.020	0.040	0.890	0.007	0.005
IKL knee extension torque (180 °/s)	0.571	-0.029	0.084	0.379	-0.046	0.212
IKR knee flexion torque (30 °/s)	0.675	-0.022	0.048	0.556	-0.031	0.096
IKL knee flexion torque (30 °/s)	0.806	-0.013	0.017	0.696	-0.020	0.040
IKR knee flexion torque (90 °/s)	0.995	0.001	0.001	0.862	-0.009	0.008
IKL knee flexion torque (90 °/s)	0.859	-0.009	0.008	0.730	-0.018	0.032
IKR knee flexion torque (180 °/s)	0.537	0.032	0.102	0.618	0.026	0.068
IKL knee flexion torque (180 °/s)	0.619	0.026	0.068	0.696	0.020	0.040
IKR elbow extension torque (30 °/s)	0.967	-0.002	0.000	0.750	-0.017	0.029
IKL elbow extension torque(30 °/s)	0.894	-0.007	0.005	0.775	-0.015	0.023
IKR elbow extension torque (90 °/s)	0.986	-0.001	0.000	0.955	-0.003	0.001
IKL elbow extension torque (90 °/s)	0.869	0.009	0.008	0.947	0.003	0.001
IKR elbow extension torque(180 °/s)	0.407	0.043	0.185	0.433	0.041	0.168
IKL elbow extension torque (180 °/s)	0.429	0.041	0.168	0.457	0.039	0.152
IKR elbow flexion torque (30 °/s)	0.924	0.005	0.003	0.895	-0.007	0.005
IKL elbow flexion torque (30 °/s)	0.984	0.001	0.000	0.804	-0.013	0.017
IKR elbow flexion torque (90 °/s)	0.968	-0.002	0.000	0.749	-0.017	0.029
IKL elbow flexion torque (90 °/s)	0.955	-0.003	0.001	0.734	-0.018	0.032
IKR elbow flexion torque (180 °/s)	0.628	-0.025	0.063	0.550	-0.031	0.096
IKL elbow flexion torque (180 °/s)	0.841	-0.010	0.010	0.686	-0.021	0.044

3.4 Subgroup copy number associations with quantitative variables.

The subgroup corrected correlations of MYLK1 (Table 5) indicates that there are differences between the subgroups therefore, all of the subgroups were individually examined for correlations between MYLK1 copy number and phenotypes associated with strength and power performance. An increased copy number in STP is associated with increases in the jump performance, handgrip, isokinetic test performance and sprinting performance. However, after the BH correction only static jump and isokinetic performance remain significant (Table 7). The strength of the correlations is stronger in STP compared to when the whole cohort is examined together as seen by a higher percentage of the variance being explained (Table 7). In the control group, only one individual has a deviation in copy number away from two therefore accurate correlations could not be performed and no associations are found in END (data presented in Appendix D). There were no significant subgroup corrected or uncorrected correlations for MYLK-AS1, therefore the subgroups were not investigated for correlations between copy number and performance in the strength and power tests.

Table 7 - Pearson's correlation coefficient results between STP for MYLK1 copy number and quantitative variables. Correlations with p<0.05 are highlighted in bold and values marked by * have a p-value lower than the acceptable alpha value according to the BH correction. kgFFM; kilogram of fat free mass IKR, isokinetic right; IKL, isokinetic left

Trait	MYLK1 CN SSP associations		
	p-value	Coefficient	R^2 (%)
Number of pull-ups	0.125	0.141	1.988
CMJ height (cm)	**0.045**	**0.184**	**3.386**
Static jump height (cm)	**0.022***	**0.214**	**4.580**
Sprint 0-10 (s/kgFFM)	**0.032**	**-0.197**	**3.881**
Sprint 10-20 (s/kgFFM)	**0.038**	**-0.192**	**3.686**
Sprint 20-30 (s/kgFFM)	0.091	-0.157	2.465
Sprint Total (s/kgFFM)	**0.043**	**-0.186**	**3.460**
Total Handgrip (kg)	**0.022**	**0.21**	**4.410**
IKR knee extension torque (30 °/s)	0.361	0.087	0.757
IKL knee extension torque (30 °/s)	0.204	0.121	1.464
IKR knee extension torque (90 °/s)	0.073	0.169	2.856
IKL knee extension torque (90 °/s)	0.044	0.19	3.610
IKR knee extension torque (180 °/s)	0.07	0.171	2.924
IKL knee extension torque (180 °/s)	**0.038**	**0.196**	**3.842**
IKR knee flexion torque (30 °/s)	0.721	0.034	0.116
IKL knee flexion torque (30 °/s)	0.6	0.05	0.250
IKR knee flexion torque (90 °/s)	0.506	0.063	0.397
IKL knee flexion torque (90 °/s)	0.496	0.065	0.423
IKR knee flexion torque (180 °/s)	0.217	0.118	1.392
IKL knee flexion torque (180 °/s)	0.325	0.094	0.884
IKR elbow extension torque (30 °/s)	**0.029**	**0.208**	**4.326**
IKL elbow extension torque (30 °/s)	0.06	0.178	3.168
IKR elbow extension torque (90 °/s)	0.223	0.118	1.392
IKL elbow extension torque (90 °/s)	0.087	0.163	2.657
IKR elbow extension torque (180 °/s)	**0.003***	**0.279**	**7.784**
IKL elbow extension torque (180 °/s)	0.065	0.175	3.063
IKR elbow flexion torque (30 °/s)	**0.046**	**0.188**	**3.534**
IKL elbow flexion torque (30 °/s)	**0.04**	**0.191**	**3.648**
IKR elbow flexion torque (90 °/s)	**0.35**	**0.089**	**0.792**
IKL elbow flexion torque (90 °/s)	0.095	0.157	2.465
IKR elbow flexion torque (180 °/s)	0.597	0.051	0.260
IKL elbow flexion torque (180 °/s)	**0.007***	**0.254**	**6.452**

The phenotypes associated with strength and power which were found to be correlated to MYLK1 copy number were examined in the overall cohort (see Appendix E) and STP (see Appendix F) for the percentage difference between copy numbers. The strongest correlations

were found for performance in isolated muscles. Overall in the cohort and in STP, the highest amount of variance explained is for the isokinetic extension of the elbow at 180 °/s. Overall individuals with three copies of MYLK1 were able to produce 38% more force than individuals with two copies, while STP with three copies are able to produce 28% more force than STP with two copies (Figure 3).

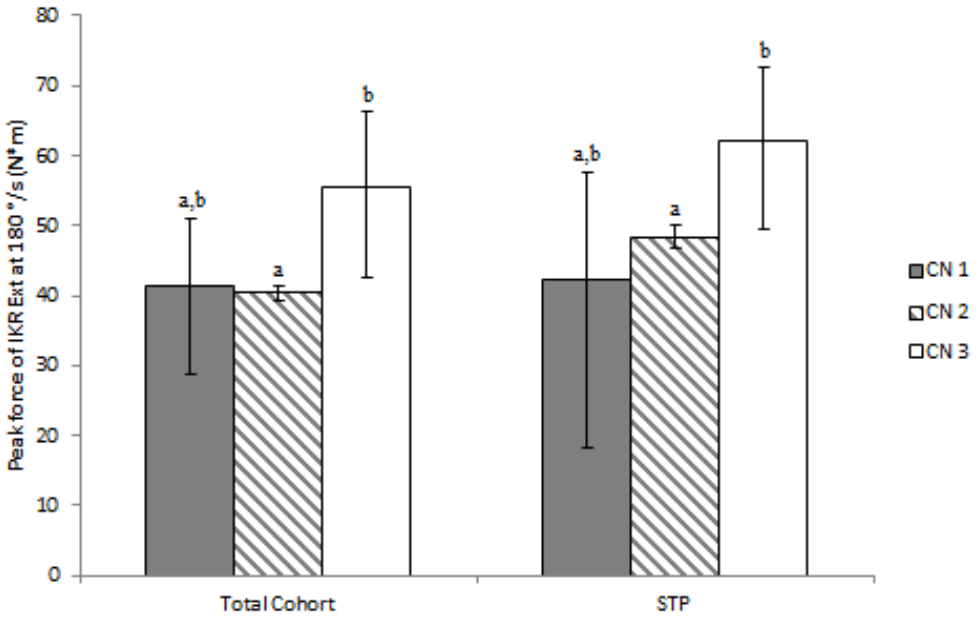

Figure 3 – Mean peak force produced (N*m) for isokinetic right arm extension of the elbow at 180 °/s for each copy number in the whole cohort and the STP alone. Copy numbers which do not share the same letter are significantly different from each other with each group. CN;copy number, IKR Ext; isokinetic right arm Extension, STP; strength and power athletes.

For performance in the whole body tasks, overall in the whole individuals with three copies were able to jump 12% and 32% higher in the static jump than individuals with two copies and one copy respectively. Static jump performance indicates that STP with two copies are able to jump 36% higher than STP with one copy and a trend existed for STP with three copies jumping higher than STP with one copy ($p=0.064$; Figure 4). Overall in the whole cohort, individuals with three copies are 11% quicker per unit of fat free mass at sprinting 30m compared to individuals with only two copies. There is a trend for STP with

three copies being quicker per unit of fat free mass than STP with one (p=0.062) or two (p=0.0791) copies (Figure 5).

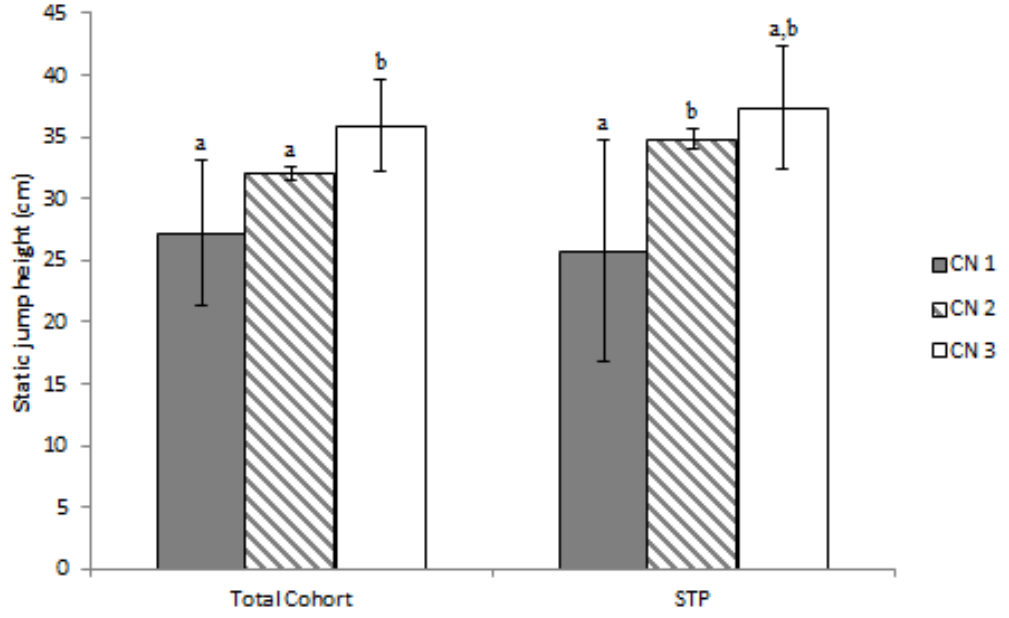

Figure 4 - Mean static jump height (cm) for each copy number in the whole cohort and the STP. Copy numbers which do not share the same letter are significantly different from each other with each group. CN;copy number, STP; strength and power athletes.

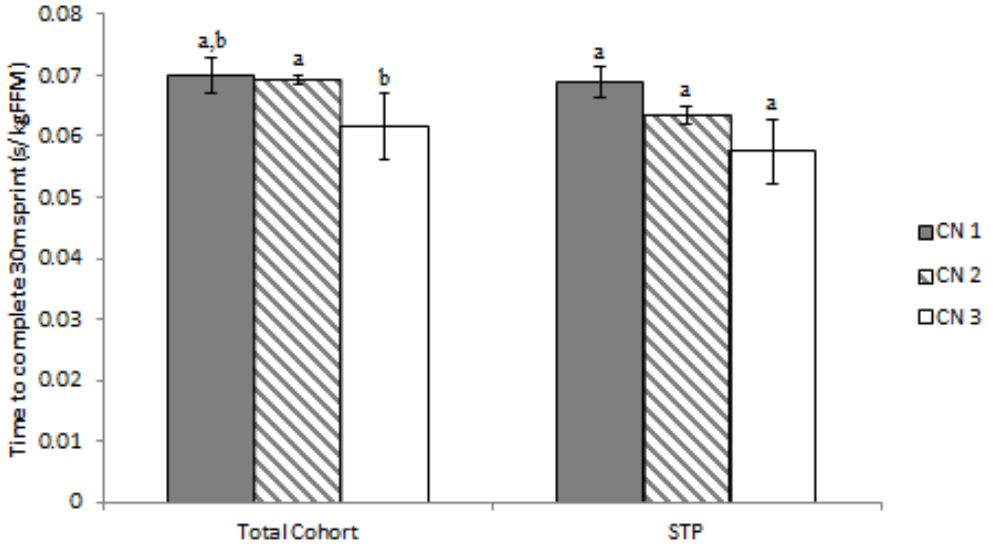

Figure 5 - Mean 30 m sprint time (presented as s/kgFFM) as for each copy number in the whole cohort and the STP alone. Copy numbers which do not share the same letter are significantly different from each other with each group. CN; copy number, kgFFM; kilograms of fat free mass, STP; strength and power athletes

Chapter 4 - Discussion

4.1 Main Findings

The current study aimed to investigate if higher MYLK1 copy number and lower MYLK-AS1 copy number is associated with athlete status and performance in physiological tests associated with strength and power performance in a cohort of athletes and non-athlete controls. We report, for the first time, increased copy number of MYLK1 is associated with athlete status and improved muscle performance. The main findings of the current study are: 1) individuals with higher MYLK1 copy number are more likely to be strength and power athletes than controls; 2) the highest level of competition achieved is not associated with MYLK1 copy number; 3) an increase in MYLK1 copy number is associated with better performance in tests indicative of strength and power, across the whole cohort and for STP alone; 4) copy number of MYLK-AS1 does not influence athlete status or performance in tasks associated with strength and power.

4.2 MYLK1

4.2.1 Copy number and athlete status

In the current study, we hypothesised that there would be an overrepresentation of individuals with a higher copy number in STP. We demonstrated that individuals with an additional copy of MYLK1 are more likely to be a STP than a control (Table 4). A deviation of copy number away from two does not confer an added chance of being an END compared to control. Taken together, these data suggest that having three copies of MYLK1 provides an advantage which predisposes individuals to becoming a STP. Although no other studies have examined

the nature of CNVs in athletes and associations with athlete status, several studies have associated CNVs with the chance of developing several diseases. There are several diseases which have been associated with the alterations of copy number, including neurological disorders (Singleton *et al.*, 2003; Sleegers *et al.*, 2006; Weiss *et al.*, 2008; Pankratz *et al.*, 2011), various forms of cancer (Yang *et al.*, 2013) and several chronic health diseases (Pollex & Hegele, 2007; Grayson *et al.*, 2010; Jeon *et al.*, 2010). There are also examples of CNVs reducing the risk of health disorders such as obesity (Falchi *et al.*, 2014) and delaying the onset of spinal muscular atrophy (Elsheikh *et al.*, 2009). The associations between CNVs and disease highlight the fact that alterations in gene copy number can be functional for a range of different traits and that consequences can be beneficial or deleterious. The mechanism by which CNVs have phenotype consequences and alter disease susceptibility remains to be fully understood but is thought to occur through a change in the expression of the affected gene (Pollex & Hegele, 2007; Grayson *et al.*, 2010; Yang *et al.*, 2013). Despite this being the first study to associate CNVs with athlete status, other forms of genetic variation, namely SNPs have been associated with athlete status.

When SNP genotype frequencies of athletes are examined there is an indication that there is a genetic predisposition to the trade-off between endurance and power capacity. Certain genotypes are over-represented in either strength or endurance athletes which indicates individuals are genetically predisposed to perform better in either strength/power or endurance-type activity (Yang *et al.*, 2003). ACTN3 is a muscle structural protein involved in the stabilization of muscle contractions and is predominantly expressed in fast twitch muscle fibres (Yang *et al.*, 2003; Moran *et al.*, 2007; Walsh *et al.*, 2008). The X allele of the R577X ACTN3 polymorphism has been shown to render the protein non-existent (Berman & North, 2010). When the genotype frequencies of R577X are examined, it is found that there is a lack

of the X allele and over-representation of the R allele in strength and power athletes including Russian power athletes (Druzhevskaya *et al.*, 2008), Greek power athletes (Papadimitriou *et al.*, 2008) Finnish sprinters (Niemi & Majamaa, 2005), American strength athletes (Roth *et al.*, 2008) and Australian sprint/power athletes (Yang *et al.*, 2003). These studies show that genetic variants which affect fast twitch fibres can be overrepresented in a cohort of power and strength athletes but not in endurance based athletes as was found in the present study.

In humans, the majority of muscles are comprised of a mixed fibre type composition, therefore it has been suggested that individuals may be predisposed to either power or endurance phenotypes (Van Damme *et al.*, 2002; Yang *et al.*, 2003). Animal models have extensively been used to determine if there is a pre-determined evolutionary trade-off in muscle capacity towards either power or endurance phenotypes. In mice a negative correlation exists between maximum power output and fatigue resistance of muscles suggesting an evolutionary trade-off (Wilson & James, 2004). In a study examining the performance of decathletes, it has been found that performance in the 1500m race is negatively correlated with the performance in the 100m, 100m hurdles, shot put and long jump (Van Damme *et al.*, 2002). These findings suggest that the performance of decathletes is either favoured towards power or endurance and indicate that in humans a trade-off between endurance and power capacity does in fact exist. Genetic variants, such as, MYLK1 copy number may influence the sport in which athletes perform as is indicated by the results of the present study.

A potential reason for the overrepresentation of higher MYLK1 copy number in the STP but not in the END is because MYLK is predominately expressed in the fast twitch type

II muscle fibres rather than slow twitch type I muscle fibres (Zhi *et al.*, 2005; Herring *et al.*, 2010; Kamm & Stull, 2011). The fibre type distributions of athletes who compete in different events have been examined in previous studies (Costill *et al.*, 1976; Zierath & Hawley, 2004). It has been found that athletes who perform endurance based events have a higher relative percentage of slow twitch fibres and sprinters have a higher percentage of fast twitch fibres (Costill *et al.*, 1976). The increased percentage of fast twitch fibres in sprinters is possibly because they are the fibres which can produce higher force and velocity, which are required to compete in short distance, high power output events (Zierath & Hawley, 2004). Therefore, it is expected that in our cohort the greatest distribution of fast twitch fibres will be found in the STP and these individuals will be the most affected by functional genetic variation in the MYLK1 gene.

4.2.2 Level of Competition

In our cohort of athletes when the highest level of competition achieved is examined for associations with MYLK1 copy number, only the regional athletes differ from the controls for increased copy number. If MYLK1 copy number was associated with the highest level of competition achieved then we would expect to find that the higher classifications of athletes (international and national standard) would differ from the lower levels of athletes (regional and other) and controls. Given the lack of associations across different competition levels, our results indicate a gain in copy number is associated with the selection of sport because of natural ability, although there might not be any further advantage gained from training.

Genetic variants for power and endurance based sports have shown increased frequencies of one allele in higher level athletes. In a study looking at 10 common genotypes which have been associated with endurance performance it is found that individuals with a higher number of 'endurance alleles' are more likely to be an elite athletes than non-elite athletes (Ahmetov *et al.*, 2009). Another study has found the frequency of the Gly482 allele of the PPARGC1A gene was 20% higher in top-level athletes compared to national level athletes (Eynon *et al.*, 2010). Associations also have been made with genotype frequencies in elite and non-elite power orientated athletes. In a cohort of elite Australian power athletes (Yang *et al.*, 2003) and a cohort of Finnish sprinters (Niemi & Majamaa, 2005), the XX genotype of ACTN3 R577X is not detected. These studies strengthen the theory that certain genotypes may predispose athletes to certain athletic disciplines since higher level athletes show an increased allele frequency compared to both controls and lower level athletes. Despite these previous findings that a range of genes have been associated with elite athlete status; MYLK1 copy number does not associate with elite athletic performance in the present study.

A potential reason for the lack of findings in the level of competition of the athletes is that both higher and lower MYLK1 copy number are rare variants and therefore found at a much lower frequency than the genetic variants examined in other studies. A gain in copy number was only found in our cohort at 2% (5% in STP) and only 12 individuals (3%) in the cohort had deviations away from a normal copy number of two (Table 3). In previous studies considering the highest level of competition achieved, the minor allele frequency has been greater than 5% in all cases and is typically around 30% (Yang *et al.*, 2003; Niemi & Majamaa, 2005; Ahmetov *et al.*, 2009; Eynon *et al.*, 2010). Another potential reason for MYLK1 copy number associating with STP but not level of competition is that the overall

ability of an athlete is determined through various genetic polymorphisms and environmental factors (Pitsiladis *et al.,* 2013). Environmental factors such as opportunity, economic possibility and social support are all involved in the determination of elite level athletes (Lucía *et al.,* 2010). The sport in which the athlete competes also needs to be considered for the level of competition achieved; regional athletes of one sport may actually be a better overall calibre athlete than a national level athlete of a different sport but competition for places in the national squad may be the limiting factor. Therefore highest level of competition achieved may not be the best measure for the effect of the genetic variants because it may not be representative of overall ability.

4.2.3 Quantitative data

The inclusion of the quantitative phenotype data strengthens the study because it allows the investigation of the impact of copy number on different aspects of performance which provides more conclusive findings than the qualitative nature of athlete status and level of competition data. Positive correlations were found between MYLK1 copy number and measures of strength and power in both isolated muscles and dynamic movements. The correlations were absent or weakened after correction for subgroup. Therefore, the differences between groups explain some of the phenotypic consequences between copy number associations. Between-group differences are not surprising because of the different training strategies required for STP and END and a lack of training performed by the controls. Each of the athletic subgroups was examined individually for the correlations between copy number and quantitative traits. Only one of the controls had a deviation away from a copy number of two. Therefore, correlations were not performed within this group. Several correlations were found in the STP for both isolated and dynamic muscle

performance (Figure 3-5), whereas no associations were found in END (Appendix D). The correlations in STP were expected because of the overrepresentation of STP with three copies of MYLK1. The existence of correlations in STP and not END suggest that it may be the type II muscle fibres which are affected or that the effect of the CNV is increased through training specific to STP.

Different polymorphisms have been associated with different aspects of performance, such as response to training or maximal force production (Kostek *et al.*, 2010; Ahmetov & Fedotovskaya, 2012). Previous MYLK1 SNP studies have identified a potential role of MYLK1 variation in the determination of baseline strength levels (Clarkson *et al.*, 2005) which could suggest that MYLK1 may make individuals stronger and therefore more likely to perform sports related to power rather than endurance. The previous associations with MYLK1 have only been examined in isolated muscle groups therefore the effect on physiological performance is yet unknown (Clarkson *et al.*, 2005). The current study associates increased copy number with both performance in isolated muscle groups and performance in whole body movements physiologically relevant to performance. The findings with phenotypes associated with isolated muscles (Figure 3), such as dynamometry measures, allow the determination of peak force, however these are not physiologically relevant measures of performance. Individuals with three copies of MYLK1 were able to jump higher than individuals with one or two copies. An advantage also is seen in sprinting performance over 30m where the individuals with three copies are quicker per unit of fat free mass than the individuals with only two copies. MYLK1 copy number is shown to explain a lower percentage of the variance in these whole body traits; however, this is expected because these whole body movements consist of the contractions of several muscles, therefore external factors have more of a chance to affect them.

Two SNPs in MYLK1 have previously been associated with muscle performance in diseased cohorts (Clarkson *et al.*, 2005; Deuster *et al.*, 2013) which supports the findings of the present study. The T allele of C49T of MYLK1 has been associated with increased baseline strength in a clinical cohort (Clarkson *et al.*, 2005). Unpublished work in this cohort of athletes reports associations with the C allele of C49T and increased force production. The differences in the allele which is associated with strength could be explained by the different cohorts that were examined. The other SNP which has been associated with muscle performance phenotypes is C37885A. The C allele of C37885A has been shown to have a protective effect on strength loss following muscle damaging exercise (Clarkson *et al.*, 2005). . Unpublished work in the present cohort has associated the C allele with increased peak force in isokinetic arm tests which supports the findings in the literature of the C allele producing an advantageous muscle performance phenotype. The SNP studies support our findings that genetic variants of MYLK1 are associated with performance in quantitative muscle performance traits.

With SNPs typically only explaining 1-2% of the variance in muscle strength between individuals, a large percentage of the variance remains unexplained (Kostek *et al.*, 2010). The CNV of MYLK1 in this study has been shown to explain at most 3% of the variance between individuals in the cohort. When the variance in the STP is examined, we demonstrate that over 7% of the performance in high speed elbow extension between individuals is explained by differences in copy number. The results indicate that CNVs may contribute more too phenotypic traits than SNPs which is expected because SNPs only affect one nucleotide; whereas CNVs can influence millions of nucleotides and thus can have more drastic effects. More studies on CNVs and athletic performance are required to confirm the associations before a statement on the size of the effect compared to SNPs can be made.

4.2.4 Mechanisms

Increased MYLK1 copy number is thought to associate with increased performance in strength and power phenotypes through an increase in the expression of the MYLK1 gene. Despite MYLK1 being expressed in skeletal muscle, skeletal muscle RLC is not a good substrate for MYLK1 because of differences in substrate properties (Kamm & Stull, 2001). In MYLK2 ablated mice the phosphorylation of the RLC still occurs which indicates there may be a redundancy mechanism between skMLCK and smMLCK despite a lack of increased MYLK1 expression as a result (Takashima, 2009). The phosphorylation of the RLC in MYLK2 ablated mice supports this because it indicates that other kinases are able to phosphorylate skeletal muscle RLC and will have the same effect on enhanced rate of force development as MYLK2(Zhi *et al.*, 2005). In cardiac tissue the smMLCK is expressed at a lower level than cmMLCK and despite differences in substrate properties, smMLCK is still capable of phosphorylating the cardiac muscle RLC. This has led to the suggestion that following the decline in cmMLCK expression smMLCK may act as a backup kinase (Josephson *et al.*, 2011). We speculate that the increased copy number of MYLK1 leads to increased expression of smMLCK which acts together with the skMLCK to produce increases in the rate of force development and force potentiation.

In humans, RLC phosphorylation has been linked to post-activation potentiation (PAP) of type II muscle fibres (Lorenz, 2011). The previous contractile history of the muscle has a large effect on the ability of the muscle to generate force and enhanced contractile response is thought to occur following high intensity contraction (Docherty & Hodgson, 2007). PAP results in increased force and rate of force development which has led to

suggestions that the molecular mechanism responsible for the enhanced contractile response is increased phosphorylation of the RLC. The phosphorylation of the RLC by MLCK occurs at a rapid rate and the dephosphorylation is a slower process, therefore it is thought that prior contractions function to prime the muscle for future contractions and subsequently increase the rate of force development and peak isometric twitch force (Zhi *et al.*, 2005). Greater PAP has been consistently observed in trained individuals compared to untrained individuals (Chiu *et al.*, 2003) and amongst individuals with a higher percentage of type II muscle fibres (Hamada *et al.*, 2000). Therefore we speculate that the increase in MYLK1 copy number will lead to increased expression of MYLK1 and subsequently increased phosphorylation of the RLC in skeletal muscle leading to elevated PAP. A potential reason that increased copy number is found more often in STP is that they are the thought to be the individuals with the greater distribution of type II muscle fibres and therefore will gain the greatest advantage from PAP.

4.3 MYLK-AS1

It was hypothesised that an increase in the MYLK-AS1 gene would lead to detriments in athletic performance. No associations were found between MYLK-AS1 copy number and athlete status or highest level of competition achieved. These findings indicate that alterations in MYLK-AS1 copy number do not confer any advantage or disadvantage to athletic performance. When the quantitative phenotype data for traits associated with strength and power are examined, it is shown that there are no positive or negative correlations with MYLK-AS1 copy number and these phenotypes. Taken together, these data indicate that alterations in MYLK-AS1 copy number is not involved in athletic performance related to strength and power. It has been shown that most CNVs have some form of effect on

expression which usually translates to alterations in the final protein content (Stranger *et al.*, 2009; Tang & Amon, 2013). In the present study, we do not find any associations between MYLK-AS1 copy number and function; however that does not mean that there is no functional consequence. The results suggest that alterations in copy number of MYLK-AS1 does not influence skeletal muscle phenotypes related to strength and power. The lack of functional consequences of CNVs of MYLK-AS1 suggests that antisense genes may not always influence the function of the associated sense genes.

Antisense genes are found across a range of different organisms (Dahary *et al.*, 2005) and at a higher frequency with genes which have multiple isoforms. It has been suggested that antisense transcription has been conserved through evolution to influence the splicing of the associated sense gene (Morrissy *et al.*, 2011). There is some evidence that antisense genes can function to alter the splicing pattern of sense transcripts (Hastings *et al.*, 1997; Annilo *et al.*, 2009). Sense and antisense transcript interaction usually affects both the expression and splicing of the sense gene; however it has been shown for a small number of genes that antisense can affect only expression or splicing alone (Morrissy *et al.*, 2011). The formation of double stranded RNA is the expected mechanism which leads to alternative splicing patterns. Antisense transcripts have been found to bind to the sense transcript to mask the splice sites of the THRA gene (Hastings *et al.*, 1997). The findings suggest that it is a consequence of sequence overlap between the sense and antisense transcripts which causes the alternative splicing. To determine the importance of sequence overlap for antisense mediated splicing, it has been examined if regions with antisense overlap are more likely to reveal antisense mediated splicing than regions without overlapping sequences. There is a 2.5 fold increase in regions with overlapping sense and antisense sequences suggesting that overlapping sequences are an important factor in antisense mediated splicing (Morrissy *et al.*,

2011). As is shown in Figure 2, there is a region of overlap between the MYLK1 and MYLK-AS1 genes. Therefore, variation in MYLK-AS1 may not affect the expression of MYLK1 rather regulate the splicing pattern; hence there may not be any phenotypic consequence on muscle strength and power of CNVs of MYLK-AS1.

The MYLK1 gene encodes at least three products, including nonmuscle MYLK1, smooth muscle MYLK1 and Telokin, which are produced through the use of different promoters rather than alternative splicing (Herring *et al.*, 2006). Alternative splicing has been shown to create multiple isoforms of nonmuscle MYLK1 transcripts (Lazar & Garcia, 1999; Hong *et al.*, 2011). If MYLK1 sense and antisense gene interactions function to only affect the splicing of nonmuscle MYLK1 then this may explain why there are no functional impacts on muscle strength and power phenotypes. It is possible that the findings with MYLK1 copy number are because of increased expression of smooth muscle MYLK1 and interactions with skeletal muscle to affect the strength and power phenotypes. The alteration of MYLK-AS1 expression may influence the splice pattern of the nonmuscle MYLK1 isoform, therefore potentially having no impact on skeletal muscle strength and power phenotypes.

4.4 Summary

4.4.1 Combined findings

Overall, we have demonstrated that higher copy number of MYLK1 leads to associations with athlete status and improved performance in physiological tests associated with strength and power. The associations between copy number and quantitative phenotypes are weakened upon correction for subgroup which suggests that the type of training

performed by the athletes has an effect on the phenotypic consequences of MYLK1 copy number alterations. On assessment of the subgroups, we find only one control participant had a deviation in copy number and that there are no associations in END. Higher copy number in the STP is shown to associate with increased performance in tests which isolate specific muscle groups and tests which rely on whole body movements. Increased MYLK-AS1 copy number was hypothesised to be detrimental to muscle performance, however we fail to find any associations between copy number in this gene and either athlete status or performance in physiological tests associated with strength and power. The lack of association with athletic performance in MYLK-AS1 suggests that the function of the antisense gene is not only the regulation of MYLK1 expression.

The current study has found that there are different phenotypic consequences between CNVs of MYLK1 and MYLK-AS1. Despite the sequence of the two genes overlapping (Figure 2), an alteration in copy number in one of the genes is not always met with alteration in copy number in the other. The different phenotypic consequences of copy number on the target genes indicates different copy number events must be detected in the two genes, otherwise the associations between athlete status and quantitative phenotypes would be identical. It was originally hypothesised that if there was a gain of MYLK1 or loss of MYLK-AS1 that this would increase MYLK1 expression and result in increased strength and power and the opposite would be true of a loss in MYLK1 or gain in MYLK-AS1 copy number. We only detect these associations with MYLK1 which suggests that the expected alteration of MYLK1 expression in skeletal muscle by deviations of copy number does not occur in MYLK-AS1. These findings suggest that the copy number variants investigated affect separate mechanisms for MYLK1 and MYLK-AS1.

It is thought that higher MYLK1 copy number exerts functional effects on athletic performance through increased expression of MYLK1. We suggest that the increased expression of MYLK1 will lead to increases in strength and power phenotypes through a compensatory mechanism between MYLK1 and MYLK2. This compensation is thought to exist because of the ubiquitous expression of MYLK1 in all types of tissue and evidence there is still RLC phosphorylation in skeletal muscle of mice with ablated MYLK2. It has been shown that phosphorylation of the RLC leads to increased rate of force production and force potentiation. The lack of association with MYLK-AS1 is thought to be explained by the fact that not all antisense genes are involved in gene expression. It has been shown that some antisense genes function to alter the splicing of the associated sense gene (Morrissy *et al.*, 2011). With evidence of alternative splicing of MYLK1 to create several isoforms of the nonmuscle MYLK1 protein (Hong *et al.*, 2011), we suggest that potentially the role of the MYLK-AS1 gene may be to control the splicing and therefore CNVs of this gene would be unlikely to affect skeletal muscle phenotypes.

The results of the present study have identified, for the first time, associations between CNVs and muscle performance in an athletic cohort. These results taken with previous SNP data on MYLK1 (Clarkson *et al.*, 2005)suggest that variation in this gene is functionally involved in the determination of muscle strength and power in both athletic and clinical populations. Replication of these findings in other cohorts and determination of the mechanisms behind these associations are required to understand the health implications of MYLK1 genotypes.

4.4.2 Strengths

The inclusion of different athletic subgroups in the GELA cohort allowed the investigation of the impact of CNVs on different types of training. The quantitative phenotype data both of whole body movements and isolated muscles allowed the examination of the functional effects these variants can have on performance. Identification of specific aspects of performance that are affected by CNVs allows the inference of potential mechanism underlying the associations between gene copy number and performance. The use of the multicopyType-it CNV Reference Probe assay instead of a single reference gene adds to the validity of the copy number determination in the current study. The Database of Genomic Variants (MacDonald *et al.*, 2014) has shown that the typical reference genes TERT and RPPH1 have whole gene spanning CNVs which could affect copy number determination in target genes (Conrad *et al.*, 2010; Xu *et al.*, 2011). If there is a CNV on one of the loci of the Type-it CNV reference assay then the copy number determination will only be altered by a fraction, therefore, the effect will be greatly reduced compared to traditional reference assays. The use of fluorescently labelled probes rather than fluorescent dyes increases the reliability of copy number determination because it has been determined that fluorescent dyes can overestimate copy number in long amplicons and AT rich sequences (Colborn *et al.*, 2008).

4.4.3 Limitations

As with all other studies there are limitations associated with the present study. It has been shown that somatic mosaicism of CNVs can occur where within different tissues there is a different copy number (Piotrowski *et al.*, 2008). The CNVs were examined in blood leucocytes, however the tissue that we are interested in is skeletal muscle. If there are

discrepancies in copy number of MYLK1 between leucocytes and skeletal muscle cells, then it would be difficult to detect associations with strength and power phenotypes. Such discrepancies could also suggest that a non-skeletal muscle mechanism causes the associated increases in power and strength. While the investigation of muscle specific mechanisms would undoubtedly offer an interesting insight to the relationship between CNV and performance, the invasive nature of muscle biopsies can lead to the loss of training, which can make it difficult to recruit athletes. Thus, collection of skeletal muscle biopsies was not possible with the current cohort. Nonetheless, by using, blood samples we were able to study a much larger cohort than would otherwise have been possible which provides greater power to detect genomic variants.

4.4.4 Future work

Future studies should aim to replicate the findings of this study in a different cohort of athletes and controls to validate the findings in MYLK1 copy number with associations to strength and power phenotypes. These studies should also examine other CNVs in genes which have been previously associated with muscle phenotypes to identify if other CNVs are associated with athletic performance. Studies should also include measures of the expression of target genes in skeletal muscle to confirm if increases in copy number are met with concurrent increases in expression. The inclusion of the expression data will provide some insight to potential mechanisms which underlie the athletic performance phenotypes. Once replication of MYLK1 and CNV of other genes have been investigated for associations with muscle performance, these polymorphisms can potentially be used to identify individuals at risk of developing muscle wasting disorders and metabolic diseases and provide them with personalised plans to maximise strength to improve quality of life.

Chapter 5 - Conclusion

In conclusion, CNVs of MYLK1 are found at a higher frequency in STP than in non-athlete controls. These variants of MYLK1 copy number were found to associate with quantitative phenotypes associated with strength and power. These associations explain more of the variance in the STP than in the cohort as a whole (~7% and ~3% respectively for variance in elbow extension peak torque). Conversely, it was thought that MYLK-AS1 would associate with decrements in performance; however, the presence of CNVs of MYLK-AS1 does not associate with any of the measures of strength and power phenotypes in this study. It is thought that the increases with MYLK1 copy number will lead to concurrent increases in the expression of MYLK1. The mechanisms underlying the functional changes in muscle strength and power phenotypes cannot be identified in the present study. The mechanism could potentially be an increased expression of smMLCK which could affect RLC phosphorylation in skeletal muscle to increase the rate of force development and peak force. The unknown function of the antisense gene in the regulation of protein-coding genes could be the reason that the alteration of MYLK-AS1 copy number does not influence strength and power phenotypes as was hypothesised. The result of the present study furthers the current knowledge of how genetic variants underpin muscle performance. Further studies are needed to confirm that variation in MYLK1 is associated with increased strength and power. If these findings can be replicated, MYLK1 could potentially be used as a target to maintain an individual's strength for improvements in athletic performance and the treatment of muscle wasting disorders.

References

Ahmetov II & Fedotovskaya ON (2012). Sports genomics: Current state of knowledge and future directions. *Cell Mol Exerc Physiol*; DOI: 10.7457/cmep.v1i1.e1.

Ahmetov II, Williams AG, Popov D V, Lyubaeva E V, Hakimullina AM, Fedotovskaya ON, Mozhayskaya IA, Vinogradova OL, Astratenkova I V, Montgomery HE & Rogozkin VA (2009). The combined impact of metabolic gene polymorphisms on elite endurance athlete status and related phenotypes. *Hum Genet* **126,** 751–761.

Almal SH & Padh H (2012). Implications of gene copy-number variation in health and diseases. *J Hum Genet* **57,** 6–13.

Annilo T, Kepp K & Laan M (2009). Natural antisense transcript of natriuretic peptide precursor A (NPPA): structural organization and modulation of NPPA expression. *BMC Mol Biol* **10,** 81.

Barreiro LB, Laval G, Quach H, Patin E & Quintana-Murci L (2008). Natural selection has driven population differentiation in modern humans. *Nat Genet* **40,** 340–345.

Benjamini Y & Hochberg Y (1995). Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* **57,** 289–300.

Berman Y & North KN (2010). A gene for speed: the emerging role of alpha-actinin-3 in muscle metabolism. *Physiology (Bethesda)* **25,** 250–259.

Bernstein BE, Birney E, Dunham I, Green ED, Gunter C & Snyder M (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74.

Beunen G & Thomis M (2006). Gene driven power athletes? Genetic variation in muscular strength and power. *Br J Sports Med* **40,** 822–823.

Boonpeng H & Yusoff K (2013). The utility of copy number variation (CNV) in studies of hypertension-related left ventricular hypertrophy (LVH): rationale, potential and challenges. *Mol Cytogenet* **6,** 8.

Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, Perusse L, Leon AS & Rao DC (1999). Familial aggregation of VO2 max response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol* **87,** 1003–1008.

Bouchard C, Daw EW, Rice T, Pérusse L, Gagnon J, Province MA, Leon AS, Rao DC, Skinner JS & Wilmore JH (1998). Familial resemblance for VO2max in the sedentary state: the HERITAGE family study. *Med Sci Sports Exerc* **30,** 252–258.

Bouchard C, Lesage R, Lortie G, Simoneau JA, Hamel P, Boulay MR, Pérusse L, Thériault G & Leblanc C (1986). Aerobic performance in brothers, dizygotic and monozygotic twins. *Med Sci Sports Exerc* **18,** 639–646.

Boulé NG, Haddad E, Kenny GP, Wells G a & Sigal RJ (2001). Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. *JAMA* **286,** 1218–1227.

Boutelle KN, Jeffery RW & French SA (2004). Predictors of vigorous exercise adoption and maintenance over four years in a community sample. *Int J Behav Nutr Phys Act* **1,** 13.

Bridges CB (1936). The BAR "gene" a duplication. *Science* **83,** 210–211.

Cantsilieris S, Baird PN & White SJ (2012). Molecular methods for genotyping complex copy number polymorphisms. *Genomics* **101,** 86–93.

Chiu LZF, Fry AC, Weiss LW, Schilling BK, Brown LE & Smith SL (2003). Postactivation potentiation response in athletic and recreationally trained individuals. *J Strength Cond Res* **17,** 671–677.

Choi J-W, Park C-S, Hwang M, Nam H-Y, Chang HS, Park SG, Han B-G, Kimm K, Kim HL, Oh B & Kim Y (2008). A common intronic variant of CXCR3 is functionally associated with gene expression levels and the polymorphic immune cell responses to stimuli. *J Allergy Clin Immunol* **122,** 1119–1126.e7.

Clarkson P, Hoffman E, Zambraski E, Gordish-Dressman H, Kearns A, Hubal M, Harmon B & Devaney JM (2005). ACTN3 and MLCK genotype associations with exertional muscle damage. *J Appl Physiol* **99,** 564–569.

Colborn JM, Byrd BD, Koita O a & Krogstad DJ (2008). Estimation of copy number using SYBR Green: confounding by AT-rich DNA and by variation in amplicon length. *Am J Trop Med Hyg* **79,** 887–892.

Conrad DF et al. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* **464,** 704–712.

Costa AM, Breitenfeld L, Silva AJ, Pereira A, Izquierdo M & Marques MC (2012). Genetic inheritance effects on endurance and muscle strength: an update. *Sports Med* **42,** 449–458.

Costill DL, Daniels J, Evans W, Fink W, Krahenbuhl G & Saltin B (1976). Skeletal muscle enzymes and fiber composition in male and female track athletes. *J Appl Physiol* **40,** 149–154.

D'haene B, Vandesompele J & Hellemans J (2010). Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* **50,** 262–270.

Dahary D, Elroy-Stein O & Sorek R (2005). Naturally occurring antisense: transcriptional leakage or real overlap? *Genome Res* **15,** 364–368.

De Moor MHM, Spector TD, Cherkas LF, Falchi M, Hottenga JJ, Boomsma DI & De Geus EJC (2007). Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Res Hum Genet* **10,** 812–820.

Deuster P a, Contreras-Sesvold CL, O'Connor FG, Campbell WW, Kenney K, Capacchione JF, Landau ME, Muldoon SM, Rushing EJ & Heled Y (2013). Genetic polymorphisms associated with exertional rhabdomyolysis. *Eur J Appl Physiol* **113,** 1997–2004.

Docherty D & Hodgson MJ (2007). The application of postactivation potentiation to elite sport. *Int J Sports Physiol Perform* **2,** 439–444.

Druzhevskaya AM, Ahmetov II, Astratenkova I V & Rogozkin VA (2008). Association of the ACTN3 R577X polymorphism with power athlete status in Russians. *Eur J Appl Physiol* **103,** 631–634.

Elsheikh B, Prior T, Zhang X, Miller R, Kolb SJ, Moore D, Bradley W, Barohn R, Bryan W, Gelinas D, Iannaccone S, Leshner R, Mendell JR, Mendoza M, Russman B, Smith S, King W & Kissel JT (2009). An analysis of disease severity based on SMN2 copy number in adults with spinal muscular atrophy. *Muscle Nerve* **40,** 652–656.

Eynon N, Meckel Y, Sagiv M, Yamin C, Amir R, Goldhammer E, Duarte JA & Oliveira J (2010). Do PPARGC1A and PPARalpha polymorphisms influence sprint or endurance phenotypes? *Scand J Med Sci Sports* **20,** e145–50.

Falchi M et al. (2014). Low copy number of the salivary amylase gene predisposes to obesity. *Nat Genet* **46,** 492–497.

Farah CS & Reinach FC (1995). The troponin complex and regulation of muscle contraction. *FASEB J* **9,** 755–767.

Faulkner J a, Larkin LM, Claflin DR & Brooks S V (2007). Age-related changes in the structure and function of skeletal muscles. *Clin Exp Pharmacol Physiol* **34,** 1091–1096.

Febrer A, Rodriguez N, Alias L & Tizzano E (2010). Measurement of muscle strength with a handheld dynamometer in patients with chronic spinal muscular atrophy. *J Rehabil Med* **42,** 228–231.

Gittings W, Huang J, Smith IC, Quadrilatero J & Vandenboom R (2011). The effect of skeletal myosin light chain kinase gene ablation on the fatigability of mouse fast muscle. *J Muscle Res Cell Motil* **31,** 337–348.

Gonzalez E et al. (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307,** 1434–1440.

Goodpaster BH, Park SW, Harris TB, Kritchevsky SB, Nevitt M, Schwartz A V, Simonsick EM, Tylavsky F a, Visser M & Newman AB (2006). The loss of skeletal muscle strength, mass, and quality in older adults: the health, aging and body composition study. *J Gerontol A Biol Sci Med Sci* **61,** 1059–1064.

Grayson BL, Smith ME, Thomas JW, Wang L, Dexheimer P, Jeffrey J, Fain PR, Nanduri P, Eisenbarth GS & Aune TM (2010). Genome-wide analysis of copy number variation in type 1 diabetes. *PLoS One* **5,** e15393.

Hai R, Pei Y-F, Shen H, Zhang L, Liu X-G, Lin Y, Ran S, Pan F, Tan L-J, Lei S-F, Yang T-L, Zhang Y, Zhu X-Z, Zhao L-J & Deng H-W (2012). Genome-wide association study of copy number variation identified gremlin1 as a candidate gene for lean body mass. *J Hum Genet* **57,** 33–37.

Hamada T, Sale DG, MacDougall JD & Tarnopolsky MA (2000). Postactivation potentiation, fiber type, and twitch contraction time in human knee extensor muscles. *J Appl Physiol* **88,** 2131–2137.

Hastings ML, Milcarek C, Martincic K, Peterson ML & Munroe SH (1997). Expression of the thyroid hormone receptor gene, erbA , in B lymphocytes: Alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. *Nucleic Acids Res* **25,** 4296–4300.

Hastings PJ, Lupski JR, Rosenberg SM & Ira G (2009). Mechanisms of change in gene copy number. *Nat Rev Genet* **10,** 551–564.

Hawkins SA, Wiswell RA & Marcell TJ (2003). Exercise and the Master Athlete--A Model of Successful Aging? *Journals Gerontol Ser A Biol Sci Med Sci* **58,** M1009–M1011.

Henrichsen CN, Chaignat E & Reymond A (2009). Copy number variants, diseases and gene expression. *Hum Mol Genet* **18,** R1–8.

Herring BP, Dixon S & Gallagher PJ (2010). Smooth muscle myosin light chain kinase expression in cardiac and skeletal muscle. *Am J Physiol Cell Physiol* **279,** 1–17.

Herring BP, El-Mounayri O, Gallagher PJ, Yin F & Zhou J (2006). Regulation of myosin light chain kinase and telokin expression in smooth muscle tissues. *Am J Physiol Cell Physiol* **291,** C817–27.

Hoebeeck J, Speleman F & Vandesompele J (2007). Real-time quantitative PCR as an alternative to Southern blot or fluorescence in situ hybridization for detection of gene copy number changes. *Methods Mol Biol* **353,** 205–226.

Hong F, Haldeman BD, Jackson D, Carter M, Baker JE & Cremo CR (2011). Biochemistry of smooth muscle myosin light chain kinase. *Arch Biochem Biophys* **510,** 135–146.

Huygens W, Thomis M a, Peeters MW, Aerssens J, Janssen R, Vlietinck RF & Beunen G (2004*a*). Linkage of myostatin pathway genes with knee strength in humans. *Physiol Genomics* **17,** 264–270.

Huygens W, Thomis M a, Peeters MW, Vlietinck RF & Beunen GP (2004*b*). Determinants and upper-limit heritabilities of skeletal muscle mass and strength. *Can J Appl Physiol* **29,** 186–200.

Iafrate a J, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW & Lee C (2004). Detection of large-scale variation in the human genome. *Nat Genet* **36,** 949–951.

Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DI, Mefford H, Ying P, Nickerson D a & Eichler EE (2009). Population

analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* **84,** 148–161.

Jeon J-P, Shim S-M, Nam H-Y, Ryu G-M, Hong E-J, Kim H-L & Han B-G (2010). Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus. *BMC Genomics* **11,** 426.

Josephson MP, Sikkink LA, Penheiter AR, Burghardt TP & Ajtai K (2011). Smooth muscle myosin light chain kinase efficiently phosphorylates serine 15 of cardiac myosin regulatory light chain. *Biochem Biophys Res Commun* **416,** 367–371.

Kamm KE & Stull JT (2001). Dedicated myosin light chain kinases with diverse cellular functions. *J Biol Chem* **276,** 4527–4530.

Kamm KE & Stull JT (2011). Signaling to myosin regulatory light chain in sarcomeres. *J Biol Chem* **286,** 9941–9947.

Kostek M, Hubal MJ & Pescatello LS (2010). The Role of Genetic Variation in Muscle Strength. *Am J Lifestyle Med* **5,** 156–170.

Kruglyak L & Nickerson D a (2001). Variation is the spice of life. *Nat Genet* **27,** 234–236.

Kumar M & Carmichael GG (1998). Antisense RNA: Function and Fate of Duplex RNA in Cells of Higher Eukaryotes. *Microbiol Mol Biol Rev* **62,** 1415–1434.

Lazar V & Garcia JG (1999). A single human myosin light chain kinase gene (MLCK; MYLK). *Genomics* **57,** 256–267.

Lobo I (2008). Copy Number Variation and Genetic Disease. *Nat Educ* **1,** 65.

Lorenz D (2011). Postactivation potentiation: an introduction. *Int J Sports Phys Ther* **6,** 234–240.

Lucía A, Morán M, Zihong H & Ruiz JR (2010). Elite athletes: are the genes the champions? *Int J Sports Physiol Perform* **5,** 98–102.

Luke a, Guo X, Adeyemo a a, Wilks R, Forrester T, Lowe W, Comuzzie a G, Martin LJ, Zhu X, Rotimi CN & Cooper RS (2001). Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. *Int J Obes Relat Metab Disord* **25,** 1034–1041.

MacDonald JR, Ziman R, Yuen RKC, Feuk L & Scherer SW (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* **42,** D986–92.

Metter EJ, Talbot L a, Schrager M & Conwit R (2002). Skeletal muscle strength as a predictor of all-cause mortality in healthy men. *J Gerontol A Biol Sci Med Sci* **57,** B359–65.

Millar DS, Horan M, Chuzhanova NA & Cooper DN (2010). Characterisation of a functional intronic polymorphism in the human growth hormone (GH1) gene. *Hum Genomics* **4,** 289–301.

Mitchell WK, Williams J, Atherton P, Larvin M, Lund J & Narici M (2012). Sarcopenia, dynapenia, and the impact of advancing age on human skeletal muscle size and strength; a quantitative review. *Front Physiol* **3,** 260.

Moran CN, Yang N, Bailey MES, Tsiokanos A, Jamurtas A, MacArthur DG, North K, Pitsiladis YP & Wilson RH (2007). Association analysis of the ACTN3 R577X polymorphism and complex quantitative body composition and performance phenotypes in adolescent Greeks. *Eur J Hum Genet* **15,** 88–93.

Morrissy AS, Griffith M & Marra MA (2011). Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res* **21,** 1203–1212.

Murthy KS (2006). Signaling for contraction and relaxation in smooth muscle of the gut. *Annu Rev Physiol* **68,** 345–374.

Niemi A-K & Majamaa K (2005). Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *Eur J Hum Genet* **13,** 965–969.

Palazzo AF & Gregory TR (2014). The case for junk DNA. *PLoS Genet* **10,** e1004351.

Pankratz N, Dumitriu A, Hetrick KN, Sun M, Latourelle JC, Wilk JB, Halter C, Doheny KF, Gusella JF, Nichols WC, Myers RH, Foroud T & DeStefano AL (2011). Copy number variation in familial Parkinson disease. *PLoS One* **6,** e20988.

Papadimitriou ID, Papadopoulos C, Kouvatsi A & Triantaphyllidis C (2008). The ACTN3 gene in elite Greek track and field athletes. *Int J Sports Med* **29,** 352–355.

Perry GH, Dominy NJ, Claw KG, Lee AS, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C & Stone AC (2008). Diet and the evolution of human amylase gene copy number variation. *Nat Genet* **39,** 1256–1260.

Pescatello LS, Devaney JM, Hubal MJ, Thompson PD & Hoffman EP (2013). Highlights from the functional single nucleotide polymorphisms associated with human muscle size and strength or FAMuSS study. *Biomed Res Int* **2013,** 643575.

Pheasant M & Mattick JS (2007). Raising the estimate of functional human sequences. *Genome Res* **17,** 1245–1253.

Piotrowski A, Bruder CEG, Andersson R, Diaz de Ståhl T, Menzel U, Sandgren J, Poplawski A, von Tell D, Crasto C, Bogdan A, Bartoszewski R, Bebok Z, Krzyzanowski M, Jankowski Z, Partridge EC, Komorowski J & Dumanski JP (2008). Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat* **29,** 1118–1124.

Pitsiladis Y, Wang G, Wolfarth B, Scott R, Fuku N, Mikami E, He Z, Fiuza-Luces C, Eynon N & Lucia A (2013). Genomics of elite sporting performance: what little we know and necessary advances. *Br J Sports Med* **47,** 550–555.

Pollex RL & Hegele R a (2007). Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation* **115,** 3130–3138.

Ramakers C, Ruijter JM, Deprez RHL & Moorman AF. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* **339,** 62–66.

Ran S, Liu Y-J, Zhang L, Pei Y, Yang T-L, Hai R, Han Y-Y, Lin Y, Tian Q & Deng H-W (2014). Genome-wide association study identified copy number variants important for appendicular lean mass. *PLoS One* **9,** e89776.

Redon R et al. (2006). Global variation in copy number in the human genome. *Nature* **444,** 444–454.

Roth SM, Ferrell RF & Hurley BF (2000). Strength training for the prevention and treatment of sarcopenia. *J Nutr Health Aging* **4,** 143–155.

Roth SM, Walsh S, Liu D, Metter EJ, Ferrucci L & Hurley BF (2008). The ACTN3 R577X nonsense allele is under-represented in elite-level strength athletes. *Eur J Hum Genet* **16,** 391–394.

Roubenoff R (2000). Sarcopenia and its implications for the elderly. *Eur J Clin Nutr* **54 Suppl 3,** S40–7.

Ruijter JM, Pfaffl MW, Zhao S, Spiess AN, Boggy G, Blom J, Rutledge RG, Sisti D, Lievens A, De Preter K, Derveaux S, Hellemans J & Vandesompele J (2013). Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. *Methods* **59,** 32–46.

Ruiz JR, Sui X, Lobelo F, Lee D-C, Morrow JR, Jackson AW, Hébert JR, Matthews CE, Sjöström M & Blair SN (2009). Muscular strength and adiposity as predictors of adulthood cancer mortality in men. *Cancer Epidemiol Biomarkers Prev* **18,** 1468–1476.

Ryan T (1974). Normal probability plots and tests for normality. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Normal+Probability+Plots+and+Tests+for+Normality#0 [Accessed December 7, 2014].

Sallis JF, Hovell MF & Richard Hofstetter C (1992). Predictors of adoption and maintenance of vigorous physical activity in men and women. *Prev Med (Baltim)* **21,** 237–251.

Sebat J et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* **305,** 525–528.

Shih RA, Belmonte PL & Zandi PP (2004). A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry* **16,** 260–283.

Silventoinen K et al. (2003). Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* **6,** 399–408.

Singleton AB et al. (2003). alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302,** 841.

Sleegers K, Brouwers N, Gijselinck I, Theuns J, Goossens D, Wauters J, Del-Favero J, Cruts M, van Duijn CM & Van Broeckhoven C (2006). APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain* **129,** 2977–2983.

Stranger BE, Forrest MS, Dunning M, Ingle CE, Thorne N, Redon R, Bird CP, Grassi A De, Tyler-smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET & Bari B (2009). Europe PMC Funders Group Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315,** 848–853.

Strub C, Alies C, Lougarre A, Ladurantie C, Czaplicki J & Fournier D (2004). Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochem* **5,** 9.

Stull JT, Kamm KE & Vandenboom R (2011). Myosin light chain kinase and the role of myosin light chain phosphorylation in skeletal muscle. *Arch Biochem Biophys* **510,** 120–128.

Takashima S (2009). Phosphorylation of myosin regulatory light chain by myosin light chain kinase, and muscle contraction. *Circ J* **73,** 208–213.

Tang Y-C & Amon A (2013). Gene copy-number alterations: a cost-benefit analysis. *Cell* **152,** 394–405.

Thomis M a, Beunen GP, Van Leemputte M, Maes HH, Blimkie CJ, Claessens a L, Marchal G, Willems E & Vlietinck RF (1998). Inheritance of static and dynamic arm strength and some of its determinants. *Acta Physiol Scand* **163,** 59–71.

Thompson PD, Moyna N, Seip R, Price T, Clarkson P, Angelopoulos T, Gordon P, Pescatello L, Visich P, Zoeller R, Devaney JM, Gordish H, Bilbie S & Hoffman EP (2004). Functional polymorphisms associated with human muscle size and strength. *Med Sci Sports Exerc* **36,** 1132–1139.

Tiainen K, Sipilä S, Kauppinen M, Kaprio J & Rantanen T (2009). Genetic and environmental effects on isometric muscle strength and leg extensor power followed up for three years among older female twins. *J Appl Physiol* **106,** 1604–1610.

Van Damme R, Wilson RS, Vanhooydonck B & Aerts P (2002). Performance constraints in decathletes. *Nature* **415,** 755–756.

Visscher PM, Hill WG & Wray NR (2008). Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* **9,** 255–266.

Walsh S, Liu D, Metter EJ, Ferrucci L & Roth SM (2008). ACTN3 genotype is associated with muscle phenotypes in women across the adult age span. *J Appl Physiol* **105,** 1486–1491.

Wang K, Li W, Glessner JT, Grant SFA, Hakonarson H & Price RA (2010). Large Copy-Number Variations Are Enriched in Cases With Moderate to Extreme Obesity. *Diabetes* **59,** 2690–2694.

Weiss LA et al. (2008). Association between Microdeletion and Microduplication at 16p11.2 and Autism. *N Engl J Med* **358,** 667–675.

Werner A (2013). Biological functions of natural antisense transcripts. *BMC Biol* **11,** 31.

Wilson RS & James RS (2004). Constraints on muscular performance: trade-offs between power output and fatigue resistance. *Proc Biol Sci* **271 Suppl ,** S222–5.

Wong KK, Ronald J, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, Macaulay C, Ng RT, Brown CJ, Eichler EE & Lam WL (2007). A Comprehensive Analysis of Common Copy-Number Variations in the Human Genome. *Am J Hum Genet* **80,** 91–104.

Xu H, Poh W-T, Sim X, Ong RT-H, Suo C, Tay W-T, Khor C-C, Seielstad M, Liu J, Aung T, Tai E-S, Wong T-Y, Chia K-S & Teo Y-Y (2011). SgD-CNV, a database for common and rare copy number variants in three Asian populations. *Hum Mutat* **32,** 1341–1349.

Yang N, MacArthur DG, Gulbin JP, Hahn AG, Beggs AH, Easteal S & North K (2003). ACTN3 genotype is associated with human elite athletic performance. *Am J Hum Genet* **73,** 627–631.

Yang T, Guo Y & Papasian CJ (2013). Copy Number Variation. In *Genetics of Bone biology and skeletal disease*, pp. 123–132. Elsevier Inc.

Yates F (1934). Contingency Tables Involving Small Numbers and the χ 2 Test. *Suppl to J R Stat Soc* **1,** 217.

Yuan JS, Burris J, Stewart NR, Mentewab A & Stewart CN (2007). Statistical tools for transgene copy number estimation based on real-time PCR. *BMC Bioinformatics* **8 Suppl 7,** S6.

Zhi G, Ryder JW, Huang J, Ding P, Chen Y, Zhao Y, Kamm KE & Stull JT (2005). Myosin light chain kinase and myosin phosphorylation effect frequency-dependent potentiation of skeletal muscle contraction. *Proc Natl Acad Sci U S A* **102,** 17519–17524.

Zierath JR & Hawley J a (2004). Skeletal muscle fiber type: influence on contractile and metabolic properties. *PLoS Biol* **2,** e348.

Appendix A – The total number of athletes and number of athletes within each level of competition who perform each of the sports.

Group	Sport	Total number	International	National	Regional	Other
SSP	Combat Sport	23	4	13	3	3
	Weightlifter	20	0	1	4	15
	Sprint Runner	13	0	5	5	3
	Sprint Swimmer	3	0	2	1	0
	Field Athlete	5	0	4	0	1
	Team sport player	53	0	18	29	6
	High-diver	1	0	1	0	0
	Motocross-rally Driver	1	0	1	0	0
	Gymnasts	7	0	5	2	0
END	Cyclists	10	1	6	1	2
	Distance Runner	49	0	29	14	6
	Distance Walker	4	3	1	0	0
	Triathlete	5	0	3	2	0
	Rower	1	1	0	0	0
	Skiers	3	1	1	1	0
	Orienteers	10	2	7	1	0
	Modern Pentathletes	2	2	0	0	0

Appendix B – Details on the physiological tests selected for phenotype determination

Handgrip strength

General upper body strength was tested with a commonly used handgrip test. A Hydraulic Hand Dynamometer (Model J00105**,** Lafayette Instrument Company, Lafayette, IN) was used to measure grip strength. Subjects were seated at the table with knees and elbow of the tested arm flexed to 90°. Each subject was permitted three trials with each hand. The dynamometer was adjusted to the individual hand size, and before each trial subjects used a towel to dry their palms to ensure a good grip. On each attempt, subjects were required to squeeze the dynamometer with maximal effort for two seconds. A rest period of 30 s was allowed between trials, and the best of the trials was recorded for each hand and totalled together to generate the total handgrip strength.

Isokinetic dynamometry. Knee and elbow flexion and extension peak torques were measured on isokinetic dynamometer (Biodex Pro3, USA) at 30, 90 and 180 °/s. Each set consisted of 3 repetitions of consecutive flexion/extensions cycles of maximal efforts throughout the range of motion. Sets always started with 30 °/s, and were followed by 90 °/s and then 180 °/s. Each of the sets was separated by a 90 s rest period, subjects started with right leg, followed by left leg, then right arm and left arm.

Pull-ups. A standard rigid gymnastic bar (2.5 cm diameter) at a height of 2.6 m from the ground was used. Participants were instructed to give their maximum efforts to perform as

many pull-ups as possible at their preferred speed but without passively hanging on the bar. The test was carefully supervised by a trained individual.

Counter movement jump and static jump. Performing counter movement jump, the participant were required to squat down quickly to 90° angle at the knee and immediately jump up as high as possible. Performing static jump, aftera 3–5 s at still squat with 90° knee angle the participant had to jump as high as possible without any additional drop in the center of gravity. Counter movement jump attempts always preceded static jumps attempts. During jumps of both modes, arms were held akimbo, and the knee angles as well as the vertical modes of the jumps were inspected visually by a single researcher. A rest period of at least 30 s between jumps was used. Jumps were performed on a portable multicomponent force platform (Kistler, type 9286AA, Switzerland), which measured the vertical component of ground reaction force at a frequency of 1000 Hz. Kistler BioWare software, with Performance module (version 3.06c), was used to collect and analyze force data during the jumps. Height (cm) of the jump was calculated using the following formula: Height = $122.625 \acute{} (T_f)^2$, where T_f = flight time (s). After practicing for 2 trials on each jump mode, three test trials of each mode were allowed, and the highest jump of each mode was taken for subsequent analysis.

30 m sprint. The sprint run was performed on a smooth hard surface. Four fixed field (without reflectors) electronic photocells (NewTest Powertimer; Oulu, Finland) for time measurement were positioned exactly at 0 m, 10 m, 20 m and 30 m at a height of 1 m, i.e., time for three consecutive 10 m sprint runs were recorded with an accuracy of 0.001 seconds. Three trials were allowed, with the recovery of 2-3 min in between. Data of the best 30 m performance were taken for the analysis.

Apendix C – Concentrations of the DNA samples used in the present study

Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)
KN001	12.79	KN043	17.35	KN082	15.69	KN123	18.13
KN002	16.24	KN044	26.64	KN083	15.44	KN124	20.03
KN003	15.53	KN045	29.65	KN085	21.15	KN125	21.78
KN004	20.7	KN046	16.89	KN086	18.04	KN126	18.83
KN005	15.93	KN047	10.84	KN087	24.39	KN127	19.28
KN006	10.5	KN048	14.9	KN088	12.08	KN128	17.9
KN007	15.3	KN049	13.17	KN089	16.31	KN129	22.6
KN008	9.44	KN050	15.297	KN090	20.21	KN130	24.43
KN009	13.54	KN051	8.6	KN091	16.2	KN131	17.17
KN010	25.66	KN052	10.92	KN092	36.62	KN132	23.66
KN011	22.44	KN053	11.8	KN093	26.74	KN133	21.41
KN012	31.61	KN054	19.46	KN094	18.34	KN134	11.61
KN013	11.4	KN055	19.01	KN095	22.04	KN135	14.03
KN015	15.33	KN056	15.23	KN096	26.61	KN137	15.17
KN017	17.93	KN057	17.09	KN097	19.58	KN138	18.36
KN020	12.29	KN058	27.29	KN098	22.75	KN139	15.97
KN021	18.36	KN059	19.64	KN099	18.16	KN140	19.84
KN022	23.23	KN060	16.53	KN100	30.52	KN142	16.14
KN023	11.4	KN061	21.92	KN101	28.97	KN143	16.26
KN024	18.76	KN062	26.91	KN102	21.42	KN145	8.48
KN025	11.66	KN063	16.97	KN103	28.99	KN146	11.53
KN026	10.46	KN064	22.78	KN104	19.92	KN147	18.08
KN027	20.97	KN065	20.65	KN105	23.86	KN148	27.04
KN028	5.71	KN066	15.23	KN106	13.72	KN149	20.8
KN029	19.24	KN067	23.79	KN107	19.45	KN150	26.25
KN030	19.42	KN068	20.13	KN108	31.78	KN152	14.35
KN031	8.71	KN069	15.85	KN110	20.64	KN153	13.64
KN032	16.83	KN070	20.62	KN111	27.57	KN154	13.7
KN033	17.41	KN071	28.32	KN112	14.06	KN155	13.93
KN034	14.96	KN072	24.34	KN113	18.52	KN156	10.21
KN035	12.44	KN074	24.16	KN114	28.95	KN157	16.55
KN036	16.64	KN075	17.79	KN115	21.33	KN158	15.61
KN037	17.55	KN076	35.7	KN117	9.62	KN159	21.67
KN038	24.71	KN077	23.39	KN118	22.03	KN160	21.44
KN039	19.67	KN078	20.42	KN119	29.31	KN161	13.27
KN040	16.37	KN079	13.99	KN120	17.21	KN162	15.88
KN041	11.18	KN080	16.99	KN121	18.34	KN163	16.23
KN042	17.08	KN081	27.55	KN122	18.6	KN164	18.63

Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)
KN165	29.25	KN208	24.47	KN258	27.103	KN303	23.41
KN166	20.22	KN209	19.95	KN259	17.888	KN304	11.566
KN167	24.97	KN210	23.16	KN260	14.453	KN305	26.293
KN168	25.08	KN211	19.9	KN262	11.531	KN306	10.408
KN169	19.13	KN212	19.19	KN263	10.428	KN307	15.833
KN170	25.21	KN213	23.37	KN264	13.77	KN308	13.516
KN171	20.06	KN215	22.13	KN265	10.405	KN309	17.873
KN172	11.12	KN216	26.78	KN266	21.706	KN310	14.699
KN173	16.43	KN217	27.02	KN267	24.038	KN311	13.649
KN174	30.52	KN218	19.35	KN268	16.28	KN313	9.191
KN176	21.04	KN219	22.00	KN269	16.76	KN314	16.789
KN177	14.43	KN220	19.80	KN270	12.075	KN315	11.396
KN178	19.19	KN223	28.65	KN271	11.383	KN316	20.078
KN179	23.38	KN225	33.54	KN272	13.566	KN317	10.955
KN180	15.13	KN226	19.04	KN273	15.844	KN318	11.531
KN181	17.04	KN227	24.82	KN274	24.709	KN319	12.623
KN182	19.31	KN229	24.28	KN276	17.014	KN320	10.465
KN183	15.76	KN230	18.02	KN277	19.277	KN321	20.263
KN184	24.46	KN231	31.16	KN278	18.072	KN322	22.458
KN185	22.48	KN232	19.38	KN279	15.885	KN323	7.078
KN186	19.79	KN233	19.06	KN280	22.321	KN324	20.201
KN187	17.61	KN234	13.25	KN281	16.925	KN326	15.706
KN188	26.82	KN236	26.64	KN282	18.923	KN327	11.383
KN189	18.33	KN237	14.32	KN284	12.457	KN328	13.172
KN190	22.37	KN238	14.239	KN285	57.31	KN329	27.863
KN191	26.47	KN239	17.528	KN286	12.117	KN330	12.629
KN192	30.87	KN240	14.387	KN287	23.062	KN331	11.526
KN193	17.53	KN241	14.094	KN288	15.374	KN332	11.566
KN194	16.63	KN242	20.598	KN289	25.247	KN333	12.596
KN195	14.97	KN245	18.43	KN290	13.402	KN334	10.408
KN196	11.91	KN246	22.791	KN291	16.612	KN335	12.522
KN197	23.98	KN247	19.668	KN292	15.337	KN336	15.833
KN198	16.46	KN248	10.982	KN294	24.962	KN337	14.155
KN199	19.87	KN250	13.694	KN295	12.011	KN338	15.706
KN200	20.4	KN251	27.442	KN296	17.711	KN340	13.552
KN201	17.81	KN252	13.219	KN297	25.336	KN341	21.163
KN202	19.6	KN253	18.62	KN298	18.575	KN342	14.751
KN203	19.09	KN254	15.337	KN299	18.669	KN343	23.611
KN205	20.02	KN255	17.422	KN300	27.382	KN344	18.112
KN206	16.13	KN256	15.287	KN301	13.727	KN345	11.566
KN207	14.16	KN257	18.852	KN302	12.623	KN346	9.31

Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)	Identifier	Conc. (ng/µL)
KN347	21.363	KN371	24.023	KN400	10.344	KN429	15.194
KN348	15.788	KN372	28.48	KN401	12.557	KN430	10.889
KN349	7.098	KN373	30.507	KN402	17.842	KN431	10.982
KN350	17.625	KN374	10.282	KN403	15.706	KN433	33.047
KN351	21.122	KN376	22.708	KN405	24.962	KN434	17.525
KN352	20.598	KN377	22.747	KN407	16.473	KN435	10.955
KN353	21.706	KN378	17.525	KN409	18.727	KN436	13.066
KN354	19.668	KN380	16.358	KN410	15.334	KN437	10.919
KN355	10.813	KN381	10.8225	KN412	18.51	KN438	12.011
KN356	35.7955	KN382	28.54	KN414	57.31	KN439	8.666
KN357	29.5765	KN383	20.93	KN415	34.662	KN440	13.552
KN358	29.143	KN384	14.239	KN416	25.477	KN441	14.652
KN359	46.007	KN385	17.528	KN417	14.652	KN442	10.38
KN360	26.7155	KN386	10.889	KN418	21.307	KN443	13.727
KN361	13.5285	KN387	10.955	KN419	15.924	KN445	15.973
KN362	20.0665	KN388	12.976	KN421	15.973	KN446	32.311
KN363	12.0495	KN391	10.299	KN422	22.453	KN447	13.694
KN365	25.844	KN394	24.585	KN423	25.745	KN448	12.11
KN366	25.1945	KN395	11.531	KN424	22.321	KN449	14.741
KN367	39.264	KN396	18.112	KN425	12.557	KN450	13.516
KN368	13.5295	KN397	15.973	KN426	16.76	KN451	15.706
KN369	12.051	KN398	10.405	KN427	31.954	KN452	13.611
KN370	26.132	KN399	12.599	KN428	11.925		

Appendix D - Pearson's correlation coefficient results between END for MYLK1 copy number and quantitative variables. IKR; isokinetic right, IKL; isokinetic left

Trait	MYLK1 Copy number END associations		
	p-value	Coefficient	R^2 (%)
Number of pull-ups	0.626	0.055	0.30
CMJ height (cm)	0.399	-0.095	0.90
Static jump height (cm)	0.783	0.031	0.10
Sprint 0-10 (s/kgFFM)	0.979	0.003	0.00
Sprint 10-20 (s/kgFFM)	0.789	-0.031	0.09
Sprint 20-30 (s/kgFFM)	0.676	-0.048	0.23
Sprint Total (s/kgFFM)	0.846	-0.022	0.05
IKR knee flexion torque (30 °/s)	0.361	0.108	1.16
IKL knee flexion torque (30 °/s)	0.686	0.048	0.23
IKR knee flexion torque (90 °/s)	0.380	0.104	1.07
IKL knee flexion torque (90 °/s)	0.708	0.044	0.20
IKR knee flexion torque (180 °/s)	0.596	0.063	0.39
IKL knee flexion torque (180 °/s)	0.727	0.041	0.17
IKR elbow extension torque (30 °/s)	0.737	0.040	0.16
IKL elbow extension torque (30 °/s)	0.451	0.089	0.79
IKR elbow extension torque (90 °/s)	0.537	-0.073	0.53
IKL elbow extension torque (90 °/s)	0.509	0.078	0.61
IKR elbow extension torque (180 °/s)	0.731	-0.041	0.16
IKL elbow extension torque (180 °/s)	0.766	0.035	0.12
IKR elbow flexion torque (30 °/s)	0.704	0.045	0.20
IKL elbow flexion torque (30 °/s)	0.805	-0.029	0.09
IKR elbow flexion torque (90 °/s)	0.542	-0.072	0.52
IKL elbow flexion torque (90 °/s)	0.366	-0.107	1.14
IKR elbow flexion torque (180 °/s)	0.701	-0.045	0.21
IKL elbow flexion torque (180 °/s)	0.562	-0.068	0.47

Appendix E – Increases in performance per copy number in the whole cohort for quantitative phenotypes which displayed significant correlations after BH corrections. IKR: isokinetic right, IKL; isokinetic left.

Trait	Copy number test	p-value	Percentage Change (%)
Static jump height (cm)	2 vs 1	0.248	17.98
	3 vs 2	**0.036**	**11.81**
	3 vs 1	**0.041**	**31.92**
Sprint 0-10 (s/kgFFM)	2 vs 1	0.315	-2.41
	3 vs 2	**0.015**	**-10.04**
	3 vs 1	0.113	-12.21
Sprint 10-20 (s/kgFFM)	2 vs 1	0.722	-1.16
	3 vs 2	**0.008**	**-11.54**
	3 vs 1	0.139	-12.52
Sprint Total (s/kgFFM)	2 vs 1	0.661	-1.11
	3 vs 2	**0.010**	**-10.97**
	3 vs 1	0.123	-11.96
IKL knee extension torque (90 °/s)	2 vs 1	0.447	13.97
	3 vs 2	**0.016**	**14.97**
	3 vs 1	0.148	31.04
IKR elbow extension torque (30 °/s)	2 vs 1	0.994	0.18
	3 vs 2	**0.002**	**28.86**
	3 vs 1	0.220	29.10
IKL elbow extension torque(30 °/s)	2 vs 1	0.676	7.43
	3 vs 2	**0.006**	**26.42**
	3 vs 1	0.110	35.81
IKR elbow extension torque(180 °/s)	2 vs 1	0.879	-2.30
	3 vs 2	**0.001**	**37.62**
	3 vs 1	0.181	34.45
IKR elbow flexion torque (30 °/s)	2 vs 1	0.949	0.90
	3 vs 2	**0.005**	**21.68**
	3 vs 1	0.195	22.78
IKL elbow flexion torque (180 °/s)	2 vs 1	0.810	-1.41
	3 vs 2	**0.002**	**27.47**
	3 vs 1	0.205	25.67

Appendix F – Increases in performance per copy number in STP for quantitative phenotypes which displayed significant correlations after BH corrections.

Trait	Test	p-value	Percentage Change (%)
Static jump height (cm)	2 vs 1	**0.012**	35.49
	3 vs 2	0.233	7.25
	3 vs 1	0.064	45.32
IKRA, extension 180d/sec., N*mLOG10	2 vs 1	0.350	14.54
	3 vs 2	**0.004**	28.33
	3 vs 1	0.106	46.99
IKLA, flexion 180d/sec., N*mLOG10	2 vs 1	0.354	14.45
	3 vs 2	**0.010**	25.07
	3 vs 1	0.069	43.14