Genetic and DNA methylation markers of ageing muscle

L HE

PhD 2019

Genetic and DNA methylation markers of ageing muscle

Lingxiao HE

A thesis submitted in partial fulfilment of the requirements of
Manchester Metropolitan University
for the degree of Doctor of Philosophy

Department of Sport and Exercise Sciences
Manchester Metropolitan University
in collaboration with Katholieke Universiteit Leuven

2019

Funding

This study was funded by the European Commission through the Erasmus Mundus ‘Move-AGE’ joint doctorate program and the Research Foundation – Flanders (FWO).

Table of Contents

Abstract

Inter-individual variation can be found in muscle mass and strength during the ageing process, and in muscular adaptations to exercise. These inter-individual differences are related to genetic and DNA methylation factors. Therefore, the objective of this thesis is to explore the role of underlying genetic polymorphisms and DNA methylation with muscle strength and mass in an ageing population. This thesis first evaluated the overall genetic association with changes in muscle mass and strength among older adults (n = 200, 60–83 yrs) who received a one year of training and were reassessed one year after the cessation of training. The genetic profile was represented as a data-driven genetic predisposition score (GPS), which was calculated based on muscle-related genetic variants selected from 170 candidates through stepwise regression. The results showed that a data-driven GPS explained 0.7% of the variance in skeletal muscle mass (SMM) and 3.2% of the variance in knee strength at baseline level, 14% of the variance in SMM and 27% of the variance in knee strength after the training, and 27–37% of the variance in the loss of muscle mass and strength after the one-year cessation of training. The thesis further compared differences in blood sample methylation patterns between sarcopenic and non-sarcopenic women (n = 48, 65–80 yrs). 6,258 differentially methylated CpGs (dmCpGs) that had different methylation levels ($p < 0.01$) between the sarcopenic and non-sarcopenic groups were identified. Genes containing these dmCpGs were involved in multiple biological pathways that were related to muscle function, actin cytoskeleton regulation and energy metabolism. A DNA methylation profile score was calculated as a weighted sum of methylation levels of sarcopenia-driven CpG sites (MSSAR, based on sarcopenia-related lasso logistic regression) and the MSSAR was negatively associated with vastus lateralis size, elbow and knee strength, and explained 10.1%, 35.5% and 40.1% of the variance, respectively. In conclusion, this thesis shows that both genetic sequence architecture and DNA methylation play a role in explaining the inter-individual differences in muscle mass and strength in older adults. An individual with a more favourable genetic profile might have not only greater baseline muscle strength, but also a higher probability to respond well to training and a smaller muscular loss after quitting the training. This study provided new insights in how the methylation status differ between weak older women compared to generally age-matched muscularly fit women. Using easily accessible blood samples, individuals at high risk of sarcopenia might be identified based on their methylation profile.

List of abbreviations

5mC	Cytosine with a methyl group at the 5' position
ACE	Angiotensin converting enzyme
$ACSA_{VL}$	Vastus lateralis anatomical cross-sectional area (paper 4)
ACSM	American College of Sports Medicine
ACTN3	Alpha-actinin-3
AIC	Akaike's information criterion
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BIA	Bioelectrical impedance analysis
BMI	Body mass index
CAD	Coronary artery disease
CCRCC	Clear cell renal cell carcinoma
CNTF	Ciliary neurotrophic factor
CON	Control group (paper 1)
CpG	Cytosin-phosphate-guanine
dmCpG	differentially methylated CpG
DNMT3a	DNA (cytosine-5)-methyltransferase 3A
DNMT3b	DNA (cytosine-5)-methyltransferase 3B

DNMT3L	DNA (cytosine-5)-methyltransferase 3-like
DZ	Dizygotic
eQTL	expression quantitative trait loci
EWAS	Epigenome-wide association scans
EWGSOP	European Working Group on Sarcopenia in Older People
FDR	False discovery rate
FIT	Fitness group (paper 1 and 2)
FTO	Alpha-ketoglutarate dependent dioxygenase
GH	Growth hormone
GLM	General linear model
GO	Gene ontology
GPS	Genetic predisposition score
GPS_{SNP}	GPS calculated from seven muscle-related SNPs (paper 4)
GWAS	Genome wide association study
HGS	Hand grip strength
HIF1A	Hypoxia-inducible factor 1-alpha
ICC	Intraclass correlation coefficient
IGF	Insulin-like growth factor
IGFBP3	Insulin-like growth factor-binding protein 3
KEGG	Kyoto encyclopedia of genes and genomes

LASSO	Least absolute shrinkage and selection operator
LBM	Lean body mass
METTL21C	Methyltransferase like 21C
MIF	Migration inhibitory factor
MMU	Manchester Metropolitan University
MS_{SAR}	DNA methylation score calculated from sarcopenia-driven CpG sites (paper 4)
MS_{SNP}	DNA methylation score calculated from CpG sites located in muscle-related genes (paper 4)
MSTN	Myostatin
MVC_{EF}	Maximum isometric elbow flexion torques at an elbow flexion angle of 60° (paper 4)
MVC_{KE}	Maximum isometric knee extension torques at a knee flexion angle of 60° (paper 4)
MWAS	Methylome-wide association study
MZ	Monozygotic
PKB	Protein kinase B
PT_{IK240}	Peak isokinetic knee extensor strength under a movement speed of 240°/s (paper 2)
PT_{IK60}	Peak isokinetic knee extensor strength under a movement speed of 60°/s (paper 2)
PT_{IM60}	Peak isometric knee extensor strength at a knee flexion angle of 60° (paper 1 and 2)
PV_{IT20}	Peak velocity of isotonic knee extension movements with a load of 20% of the peak isometric strength (paper 2)
RF	Random forest

RM	Repetition maximum
RSS	Residual sum of squares
SD	Standard deviation
SMI	Skeletal muscle index
SMM	Skeletal muscle mass
SNP	Single nucleotide polymorphism
SPARC	Secreted protein acidic and rich in cysteine
TGS	Total genotype score
THK_{BB}	Biceps brachii thickness (paper 4)
VDR	Vitamin D receptor
VO_{2max}	Maximal O_2 uptake
WBV	Whole-body vibration group (paper 1 and 2)

PART 1

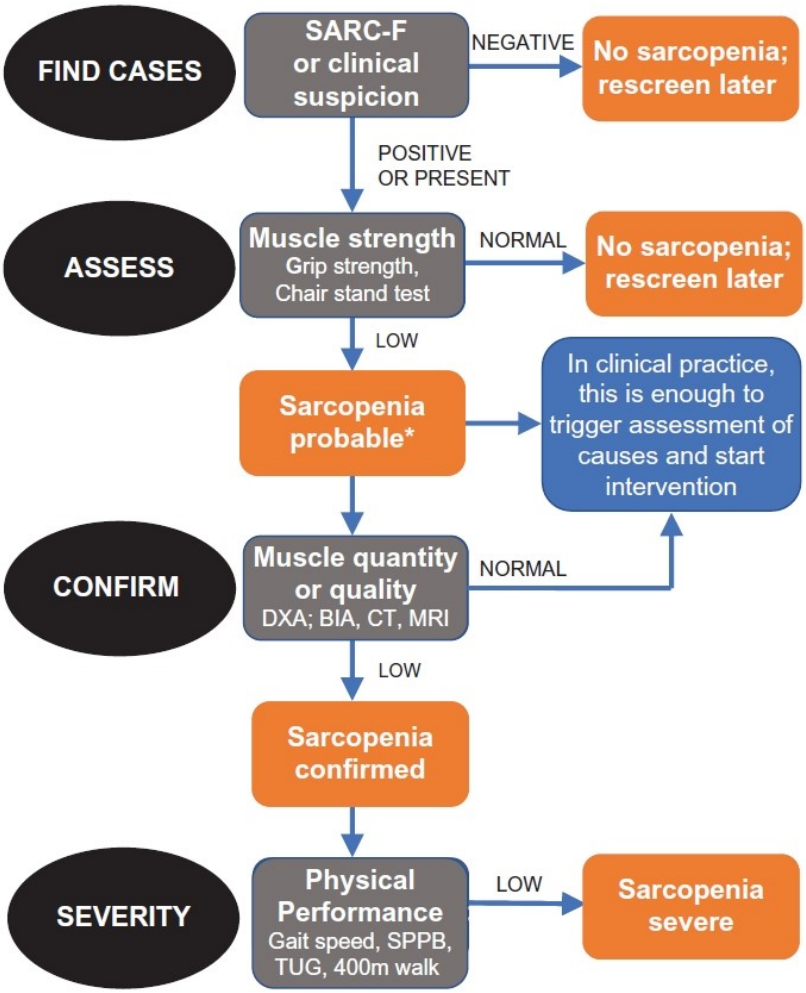
GENERAL INTRODUCTION

1. Ageing-related muscle degeneration

1.1. Demographics and ageing-related muscle degeneration

Decreased muscle mass and muscle strength are two predominant changes during ageing. It has been reported that a noticeable atrophy of skeletal muscle can be observed after an age of 40.[1] The degree of muscle degeneration aggravates with the increase of age. It is estimated that after an age of 50, muscle strength decreases at an annual rate of 1.5% while the decline of muscle mass is 1–2%.[2] Frontera et al. conducted a 12-yr follow up study on older men (mean age 65 yrs) and found a 1.7–2.5% annual decrease in elbow and knee strength, and a 1–1.3% loss in thigh muscle size.[3] Goodpaster et al. reported an annual leg strength decline of 2.6–4.1% and leg lean mass loss of 1% among older people aged 70–79 yrs.[4] Charlier et al.[5] analysed different types of knee strength (isometric, isotonic, isokinetic and endurance strength) among individuals aged between 18 and 78 years, and found that age accounted for 5–24% of the variance in knee strength in both men and women. Compared to muscle mass, muscle strength is a stronger predictive parameter for functional ability and living quality.[6] Meta-analyses on older adults aged above 65 yrs have shown that BMI (greater than 30) and low muscle strength are closely associated with functional decline while low muscle mass fails to show any significance.[7] Schaap et al.[8] analysed associations of muscle mass, grip strength and gait speed with three-year longitudinal data of falling incidents among 498 older adults (aged above 65 yrs), and reported that only low grip strength was associated with the incidence of recurrent falling. Specifically, the concept of “sarcopenia” was introduced in 2010 by the European Working Group on Sarcopenia in Older People (EWGSOP).[9] The term “sarcopenia” refers to a syndrome which is characterised by progressive loss of skeletal muscle mass and muscle function (muscle strength or performance).[9] Factors such as age, endocrine, neuro-degeneration, disuse and malnutrition are closely related to sarcopenia.[9] The second meeting of EWGSOP suggested a Find-Assess-Confirm-Severity (F-A-C-S) algorithm (I-Figure 1) which identifies sarcopenia through a criterion sequence of (1) low muscle strength (defined as probable sarcopenia), (2) low muscle mass (defined as diagnosed sarcopenia), and (3) low physical performance (defined as severe sarcopenia).[6] However, the cut-off point of each muscular parameter for sarcopenia identification is still under debate. Some studies use a 2-standard deviation (SD) below the mean of young adults as a cut-off point,[10–12] some define sarcopenia based on quartiles of the study group[13–15], others also apply statistical analysis such

as the likelihood ratio for sarcopenia classification.[16] Only recently, cut-off points for sarcopenia tests are advised by the EWGSOP with a grip strength cut-off value of 27 kg for men and 16 kg for women, a total skeletal mass value of 20 kg for men and 15 kg for women, and a gait speed of 0.8 m/s.[6] Based on the suggested cut-off points, the prevalence of sarcopenia among cummunity-dwelling Japanese older men and women (65+ yrs) is 10.1% and 7.2%, respectively.[17] While in a Gambian population (aged 40–75+ yrs), the prevalence of sarcopenia reaches 19% in men and 10% in women.[18] Meta-analysis based on 35 studies (older people aged 60+ yrs) showed that the overall estimated prevalence of sarcopenia was the same (10%) in both men and women while non-Asian had higher prevalence than Asian in both genders (19% vs 10% in men; 20% vs 11% in women).[19]



I-Figure 1. The FACS algorithm suggested by EWGSOP2 for the identification and severity quantification of sarcopenia. The identification of sarcopenia follows a checking sequence of muscle strength, muscle quantity/quality and physical performance. Adapted from Cruz-Jentoft et al., 2018.

Population ageing has become a universal phenomenon. In Europe, 25% of the population was aged 60+ in 2017 and the proportion is estimated to reach 34% in 2050.[20] Meanwhile, it is estimated that 5–13% of older people aged 60–70 years are affected by sarcopenia, and the proportion increases to 11–50% for those aged 80 or above.[2] Since muscle degeneration can greatly limit physical function,[21] and increase risk of falls[22] and mortality rate,[23] more older adults might be at the threat of reduced quality of life[11] and raised healthcare expenditures[28] in the future. Notably, the muscle degeneration rate varies between older men and women. Kyle et al. reported a 8.9 kg (14.8%) decrease in fat free mass from middle-aged to older (85+ yrs) men and a 6.2 kg (14.3%) decrease from middle-aged to older (85+ yrs) women.[25] Gallagher et al.[26] reported that the annual loss in appendicular muscle mass was approximately 0.8 kg (3.3%) in older men and 0.4 kg (2.2%) in older women over a 7-year period. The loss of muscle strength with aging is also found to be greater and faster in men. Goodpaster et al.[4] examined changes in muscle strength among older adults over a three-year period and reported that older men lost twice as much isometric knee extensor strength as women over a 3-year period. Despite the greater losses of muscle mass and strength among older men, older women might be more vulnerable during ageing. Firstly, older women have lower levels of muscle mass and strength. In general, men have 11 kg more skeletal muscle than women and women have 40% and 30% less muscle in the upper and lower body, respectively.[27] Secondly, the dramatic drop of sex hormones (e.g. estrogen and progesterone) after menopause is closely connected with decreased bone mineral density, muscle strength and lean body mass.[28,29] Baumgartner et al.[10] reported that healthy postmenopausal women (aged less than 74 yrs) had much lower appendicular SMI than men with a similar age in both Hispanics and non-Hispanic white populations. **Consequently, older women will have a lower physical performance**[30] **and a higher injury risk**.[31]

1.2. Mechanisms underlying ageing-related muscle degeneration

Many factors can contribute to ageing-related muscle degeneration. Firstly, at the neuromuscular level, myofiber denervation and loss of motor units are found with ageing.[1] Aare et al. reported that denervated myofibers were 35–50% smaller than innervated fibers in rat ageing muscle.[32] The same team later found that the accumulation of denervated myofibers in ageing muscle was due to failed reinnervation and suppression of neurotrophin response.[33] Secondly, the quantity and type of myofibers also change during ageing. Lexell et al.[34] reported a quadratic

correlation between age and the amount of myofibers, and an average reduction of 39% in myofibers from 20 to 80 yrs. A 10–40% reduction in the size of type II myofibers, a fast-twitch myofiber that exerts more power and fatigues faster than type I fibers during contraction, was also found in older participants when compared with young controls.[35] Meanwhile, a preferential atrophy of type II fibers with an elevated type I/type II fiber ratio was reported in ageing muscle.[36,37] Thirdly, at the level of muscle metabolism and function, Rooyackers et al. found decreased mitochondrial enzyme activity in ageing muscle and 12% lower mitochondrial protein synthesis rates in older participants than young controls.[38] Porter et al. demonstrated declined mitochondrial respiratory capacity and coupling efficiency with age.[39] Reduced synthesis rates of myosin heavy chain and mitochondrial protein with progressive declines in hormones (e.g. growth hormone [GH] and plasma insulin-like growth factor-I [IGF-I]) were found in ageing muscle by Proctor et al.[40] Moreover, muscle properties also change with age. For instance, older women have shown less passive dorsiflexion angle and higher passive elastic stiffness in the calf muscle-tendon unit than young controls.[41]

1.3. Non-invasive interventions for muscle degeneration

Regular exercise is well known as an effective non-invasive intervention in slowing down muscle degeneration. Resistance training and fitness training combining both resistance and aerobic exercise have been found to improve muscle performance in older adults. A meta-analysis based on 1,079 healthy participants aged 50+ showed that muscle strength (leg press, chest press, knee extension and latissimus pull down) increased by 9.8–31.6 kg (24–33%) after resistance training with a training period from 6 to 52 weeks and a training intensity of 40–85% of 1 repetition maximum (RM).[42] Healthy older men experienced a 7–10% improvement in muscle power and 8–15% increase in muscle quality (represented as a ratio of maximum muscle strength to muscle size) after 30-wk of combined (resistance and aerobic) fitness training.[43] Villareal et al. conducted a 26-wk exercise program on obese older subjects and found a 19% improvement in thigh muscle strength after resistance training, and an 18% increase in strength after combined resistance and aerobic training.[44] Besides conventional resistance or combined fitness training, whole-body vibration (WBV) training is also an effective intervention method in muscle training. By applying mechanical vibrations to skeletal muscles, WBV improves muscle characteristics in various aspects. The synchronization of motor units and the sensitivity of the stretch reflex are increased after WBV, leading to enhanced muscle

performance.[45] Delecluse et al. reported a 9–16% enhancement in knee extensor strength and a 7.6% improvement in muscle counter-movement jump after a 12-wk WBV training.[46] Roelants et al. even found a higher improvement in lower limb power induced by WBV (with 19% increase) than 8RM resistance training (with 13% increase) after a 24-wk training among older women.[47] Furthermore, researches have revealed that exercise with extra nutritional supplement will enhance the training effect. Phillips[48] summarised that participants had greater muscle mass and strength improvement through resistance training with additional supplementation of nutrients such as protein, creatine, β-hydroxy-β-methylbutyrate (β-HMB) and omega-3 polyunsaturated fatty acids (PUFAs). The study of Rodacki et al.[49] on older women (aged above 60 yrs) demonstrated that an addition of omega-3 during resistance training contributed to a greater improvement in muscle strength and functional performance than resistance training without any nutrient supplement. Cermak et al.[50] performed a meta-analysis on 22 studies which combined protein supplementation (> 1.2 $g \cdot kg^{-1} \cdot d^{-1}$) and resistance training (a training frequency of minimum twice per week for more than six weeks), and found that resistance training with protein supplementation led to a greater training response in leg press strength and fat free mass compared with a placebo in both young and older participants.

Muscle adaptations to training have been reported with a high degree of inter-individual variance.[51] Ahtiainen et al. studied 287 participants (aged 19–78 yrs) and reported an extensive variance in individual muscle responses to resistance training, with the change in leg muscle size ranging from -11 to 30% and in muscle strength ranging from -8 to 60% regardless of age and sex.[52] Similar inter-individual difference was also reported in elbow flexor size (ranging from -2 to 59%) and strength (ranging from -32 to 149%) in the study of Hubal et al.[53] Such inter-individual variance in muscle size and strength responses to training is partially associated with genetic factors. In a study on a 10-wk strength training among male young twins, Thomis et al.[54] reported that the interaction between genotype and training accounted for 21% and 9% of the variance in elbow flexor 1 RM strength and isometric elbow strength (at a 110° elbow flexion). The FAMuSS study conducted by Thompson et al., who examined genetic correlations with training-induced muscle size and strength changes among one thousand adults (aged 18–40 yrs) over a period of 12 weeks, aimed to identify genetic variants associated with muscle size and strength, and muscle responses to resistance training.[55] Analyses

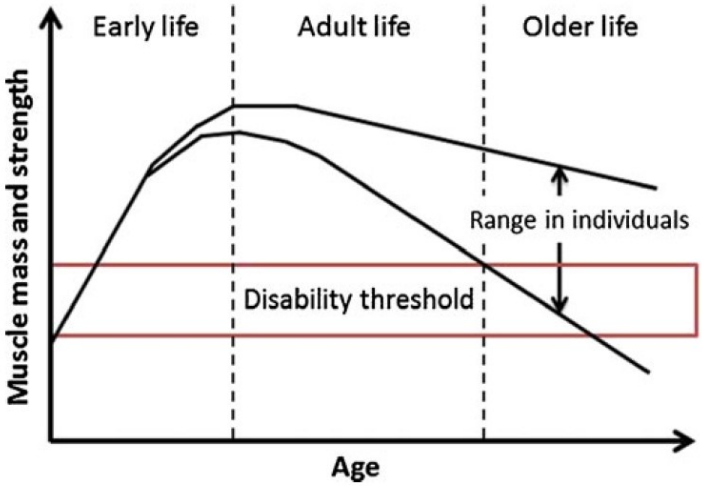
on the FAMuSS study have reported 17 genes that are associated with muscle size and strength at baseline and in response to resistance training.[56] Notably, most studies on the genetic association with muscle adaptations to exercise are based on young and middle-aged participants, **the genetic association with muscular changes in older adults is rarely studied**. Moreover, all studies are focusing on training-induced muscle changes, the time-associated loss of muscle size and strength after a training intervention (detraining effect) is less well studied and **studies focusing on genetic factors contributing to the individual differences in these detraining changes are lacking**.

1.4. Genetics and ageing muscle

Inter-individual variability in muscle mass and strength can be found not only in exercise intervention, but also during the ageing process (I-Figure 2),[57] which is believed to be partly gene-related. Studies in older twins demonstrated that hand grip strength had a heritability of 30–52%.[58–60] A 3-yr follow-up study by Tiainen et al. found that genetic factors explained 58% of the variance in knee extensor strength at baseline and 56% at follow-up, and 67% of the variance in knee extensor power at baseline and 48% at follow-up.[61] A recent meta-analysis by Zempo et al.[62] on 58 measurements regarding the heritability of muscle-related phenotypes (e.g. grip strength, isometric and isotonic strength) revealed that genetic factors accounted for 49–56% of the variance in muscular phenotypes and the environmental effect on skeletal muscle strength performance increased with age.

While twin studies demonstrate a heritable component in skeletal muscle mass and strength, genotype or allelic association studies of DNA sequence variants in candidate genes between different groups (e.g. athletes vs. controls, people with high strength vs. those with low strength) have identified many genetic variants that are associated with muscle mass and strength.[63–66] For instance, Cho et al. found that older adults with the alpha-actinin-3 (*ACTN3*, rs1815739*)* T/T genotype had a significantly higher risk of sarcopenia and osteoporosis than the C allele-carrier counterparts.[67] The D allele in the angiotensin converting enzyme (*ACE*, rs4341) gene was found to be correlated with greater lean body mass (LBM), isometric and isokinetic quadriceps strength than the I allele.[68,69] Arking studied 363 community-dwelling women (aged 70–79 yrs) and reported that AA homozygote carriers in the ciliary neurotrophic factor (*CNTF*, rs1800169) gene showed lower grip strength than the G allele carriers.[70] Heffernan et al.[71] reported that the T allele in the alpha-

ketoglutarate dependent dioxygenase (*FTO*, rs9939609) A/T polymorphism was predisposing to increased LBM and was more prevalent in elite rugby players, who rely more on appendicular lean mass for success, than other rugby athletes and non-athletes. Studies on the hypoxia-inducible factor 1-alpha (*HIF1A*, rs11549465) C/T polymorphism have shown that power-oriented athletes have a higher frequency of the T allele in comparison with controls who had no competitive sport experience.[72–74] The C allele in the myostatin (*MSTN*, rs1805086) gene has been related to a lower leg strength and functional performance (e.g. vertical jump, gait and balance) in both young and older adults.[75–77] Studies on the vitamin D receptor (*VDR*, rs2228570) G/A polymorphism reveal that the G allele carriers have less quadriceps strength and a 2.17-fold higher risk for sarcopenia than the A allele carriers.[78–80]



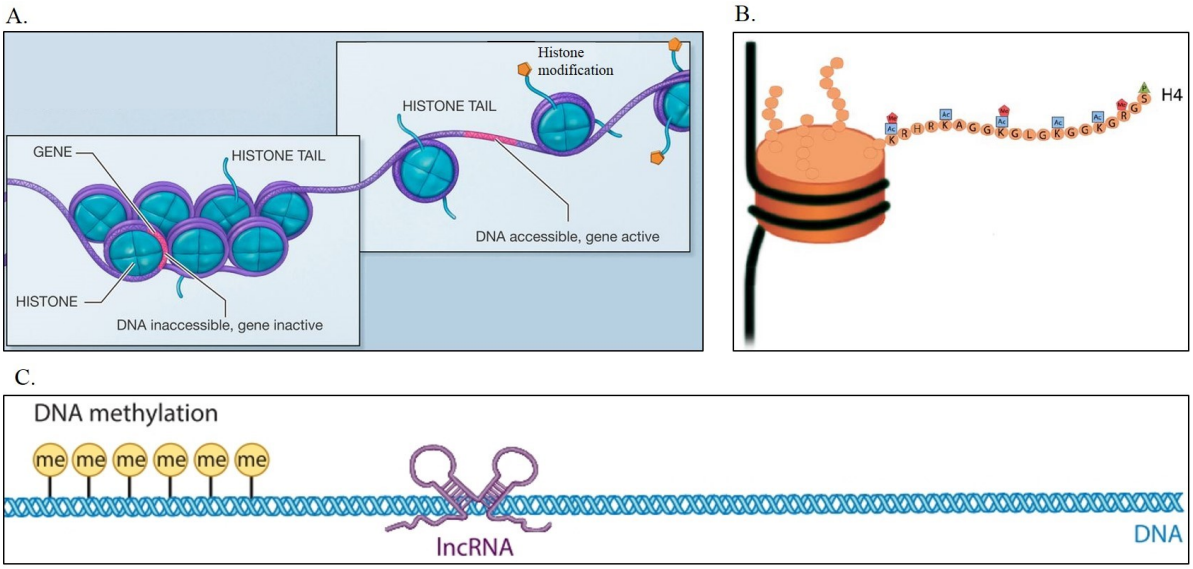
I-Figure 2. Changes in muscle mass and strength throughout the life course. Inter-individual variance gradually increases with age. Adapted from Mithal et al., 2013.

1.5. DNA methylation and ageing muscle

Besides genetic variants, the muscle can also be influenced by many environmental factors such as physical activity and nutrient intake.[81] For instance, protein intake is essential for protein synthesis and skeletal muscle function, while lower protein intake is usually found in older adults due to reduced appetite and chewing efficiency.[82] One mechanism for environmental factors to affect muscle mass and strength is through epigenetic regulation.

Epigenetic regulation includes a set of regulatory processes that modify gene expression without changing its original DNA sequence.[83] Several mechanisms such as histone modification, DNA methylation, noncoding RNA regulation and

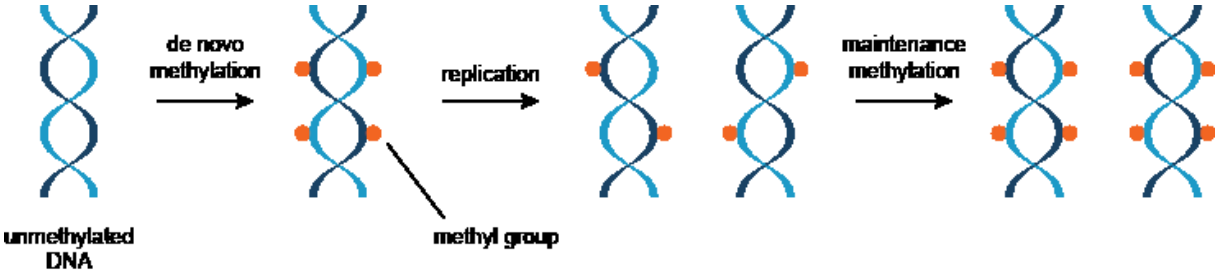
chromatin remodelling are involved in epigenetic regulation (I-Figure 3A).[84,85] Histones are the fundamental components of chromatin and are positively charged with many lysine and arginine residues. DNA, which is negatively charged due to its phosphate groups, is winded tightly around histones, and therefore, is kept from being transcribed. Covalent modifications at the N-terminal region of histone (known as the histone tail) will affect the connection between DNA strands and histones. For instance, with H3K27 acetylation, the positive charge of histones is reduced, resulting in decreased binding of DNA strands.[85] Phosphorylation can also decrease the binding while methylation on histone tails will strengthen the binding (I-Figure 3B).[86] Through these modifications, the accessibility of a DNA sequence can be regulated.



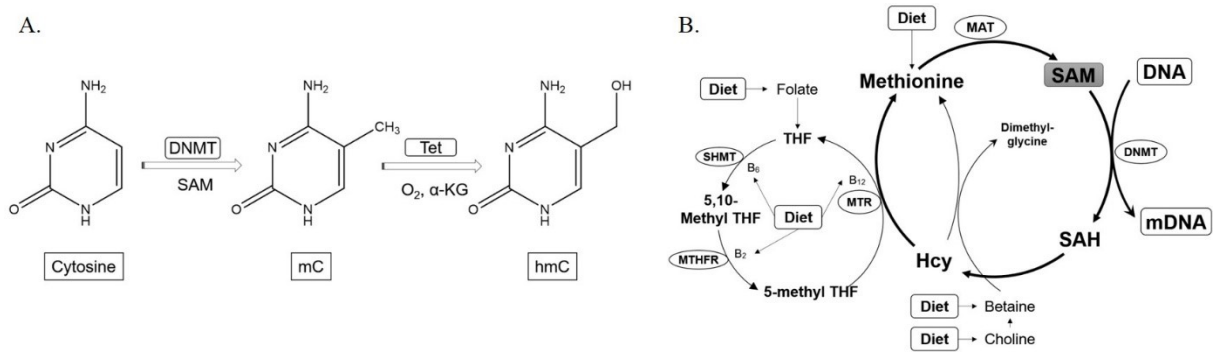
I-Figure 3. Main epigenetic mechanisms. (A) Histone modification affects the binding with DNA strands. Adapted from Brown, 2015. (B) Covalent modifications at the N-terminal region of histone H4. Covalent modifications include phosphorylation (P), acetylation (Ac), and methylation (Me). Single letter abbreviations for amino acid residues: A = alanine; G = glycine; H = histidine; K = lysine; L = leucine; R = arginine; S = serine. Adapted from Stephens, 2013. (C) DNA methylation and long noncoding RNAs (lncRNAs) regulation also affect gene transcription. Adapted from Chen, 2017.

DNA methylation, which has been extensively studied, is an epigenetic mechanism that occurs on the DNA itself (I-Figure 3C). In vertebrates, DNA methylation mainly refers to an attachment of a methyl group at the 5' position of cytosine (5mC), a process that involves DNA methyltransferases (DNMTs) for catalysis. There are three DNMTs (DNMT1, DNMT3A and DNMT3B) with different functions. DNMT3A

and DNMT3B are responsible for creating new methylation patterns (known as *de novo* methylation) and DNMT1 is involved in the maintenance of methylation patterns during cell duplication (I-Figure 4). Therefore, a DNA methylation pattern can be either a newly-formed response to a recent stimulus, or a copy of a pattern that is caused by a previous factor long time ago and is passed down from cell to cell through mitosis. A methylated cytosine (mC) can be oxidized by dioxygenases from the ten-eleven translocation (TET) family into 5-hydroxymethylcytosine (hmC), a process known as demethylation (I-Figure 5A). Moreover, DNA methylation can be affected by diet. This is because several nutrients (e.g. folate, betaine and choline) are methyl donors in the methionine cycle for the synthesis of S-Adenosylmethionine (SAM), which is an electrophilic methyl source for DNA methylation (I-Figure 5B).[87]



I-Figure 4. DNA methyltransferases. DNMT3A and DNMT3B are involved in the de novo methylation, DNMT1 participates in the maintenance of methylation. Adapted from: https://www.atdbio.com/content/56/Epigenetics



I-Figure 5. (A) Cytosine methylation and demethylation. S-Adenosylmethionine (SAM) donates methyl groups for cytosine methylation which is mediated by DNA methyltransferases (DNMTs). Methylated cytosine (mC) is demethylated into 5-hydroxymethylcytosine (hmC) via the ten-eleven translocation (TET) family. (B) Possible pathways for nutrients (folate, B vitamins, betaine and choline) to affect DNA methylation process. MAT = methionine adenosyltransferase; SAM = S-adenosyl methionine; SHMT = serine hydroxymethyltransferase; THF =

tetrahydrofolate; DNMT = DNA methyltransferase; MTR = 5-methyltetragydrofolate-homo- cysteine methyltransferase; MTHFR = methylentetrahydrofolate reductase; Hcy = homocysteine; SAH = S-adenosylhomocysteine; mDNA = methylated DNA. Adapted from Zhang, 2015.

In the human genome, DNA methylation can be found in 57–85% of cytosin-phosphate-guanine (CpG) sites.[88] CpG sites take up less than 1% of the human genome and most of those sites are scattered. Regions with a high frequency of CpG sites are defined as CpG islands, which can be found in 72% of gene promoters.[89] Although 63–81% of CpGs are heavily methylated (hypermethylated) across different tissues, the majority of CpG islands remain lowly methylated (hypomethylated).[90] The increased methylation of those CpG islands in gene promoters has been associated with the repression of gene expression.[91] The possible mechanism is that methylated CpG sites in transcriptional regulatory regions (e.g. promoters, enhancers) can inhibit the binding of transcription factors, which initiate and regulate the transcription of corresponding genes.[92] Meanwhile, the role of DNA methylation in intragenic regions remains controversial. Lorincz et al.[93] reported that intragenic DNA methylation in transcriptionally active genes could impede gene expression by reducing elongation efficiency of RNA polymerase II. Yet, a recent study by Jeziorska et al.[94] has suggested a positive association between the CpG island methylation in intragenic regions and transcriptional activity. Notably, DNA methylation patterns are tissue specific.[95] Slieker et al.[96] studied age-related DNA methylation changes in multiple tissues and found that the majority of identified differentially methylated genetic sites (85.2% of sites with increased methylation and 97.4% of sites with decreased methylation) could be found in only one of the seven tissues (brain, buccal, liver, kidney, subcutaneous fat, monocytes and T-helper cells). Therefore, methylation data obtained from one tissue can not fully represent the methylation status in other tissues. Yet, studies on ageing still identified some genetic sites with consistent methylation patterns across tissues. Horvath[97] selected 353 genetic sites for the prediction of DNA methylation age which was found to be highly correlated with chronological age across multiple tissues and even in chimpanzees. Among the differentially methylated genes across the seven tissues, Slieker et al.[96] identified the gene *ELOVL2* in all the tissues and 12 other genes in six out of the seven tissues. These findings suggest the possibility of estimating certain genetic methylation patterns in tissues (e.g. brain and muscle),

which are not easily accessible, using the methylation patterns of corresponding genes in some easily accessible tissues (e.g. blood, saliva).

The association between ageing and DNA methylation has been studied in multiple tissues. Ageing is previously believed to be characterized by a global decrease of DNA methylation (quantified as the 5mC content at CpG sites).[98] However, with the application of new technologies (e.g. bisulfite conversion and next-generation sequencing) which increase the accuracy in the 5mC content measurement, recent studies fail to observe any significant ageing-related alterations in global methylation and the expression of DNA methylation enzymes is also not associated with age.[99] Despite the inconsistent finding in ageing-related global methylation, many studies demonstrate region-specific DNA methylation differences such as increased methylation levels in gene promoters and enhancers, and methylation changes at specific CpG and CpH (H = A, C, or T) sites.[98,99] Based on ageing-related DNA methylation patterns, several CpG sites have been suggested as possible predictors of chronological age. Bocklandt et al.[100] identified 88 CpGs that were correlated with age based on saliva samples collected from male twins. They further developed an age predictive model based on the ageing-related CpGs located in the *EDARADD* and *NPTX2* genes and found that the model explained 73% of the variance in age in an independent population.[100] By comparing skeletal muscle tissue between older and young adults, Zykovich et al.[101] found 2,114 ageing-related CpGs, which were further found to be closely connected with neural control, cytoskeleton function and cell growth. The researchers also suggested 500 most significantly changed CpGs that might be used for age prediction in the future.[101]

Since strength training is beneficial to skeletal muscle, recent studies also explored the association between DNA methylation and training. Seaborne et al.[102] studied the methylation changes during a 22-wk resistance training-detraining-retraining cycle. They identified four genes (*AXIN1*, *GRIK2*, *CAMK4* and *TRAF1*) with hypomethylation and enhanced expression after resistance training and retained hypomethylation during detraining.[102] The same research team also found five genes (*UBR5*, *RPL35a*, *HEG1*, *PLA2G16* and *SETD3*), which showed increased gene expression after training, and the largest increases in hypomethylation and gene expression during the retraining.[102] The findings of Seaborne et al. suggest the existence of an epigenetic memory which might help skeletal muscle quickly adapt to a pre-encountered stimulus.[103] Turner et al.[104] further compared gene expression with methylation data after resistance training and found genes, which

were significantly enriched in both transcriptome and methylome analysis results, were involved in pathways associated with cancer, protein synthesis and actin cytoskeleton regulation. They also suggested five genes (*FLNB*, *MYH9*, *SRGAP1*, *SRGN* and *ZMIZ1*), which were hypomethylated during training and maintained hypomethylated during detraining, as being associated with epigenetic memory in skeletal muscle.[104] In another study of long term physical activity, Sailani et al.[105] found 714 hypomethylated gene promoters in older adults with regular physical activity when compared with inactive controls. Corresponding genes were associated with several energy metabolism and myogenesis pathways such as glycogen metabolism, TCA cycle, actin polymerization and oxidative stress resistance.[105] **Notably, DNA methylation patterns related to inter-individual variance in muscle mass and strength (e.g. methylation patterns between sarcopenic and non-sarcopenic subjects) have never been studied.**

1.6. Analyses in genetic and DNA methylation studies of ageing muscle.

1.6.1. Estimation of heritability

In twin studies, the degree of heritability is estimated by quantifying the contribution of an unmeasured genetic component (as a latent variable) to account for the observed variation in phenotypic traits.[106] Heritability is calculated as the ratio of genetic variance to total phenotypic variance while the genetic variance is estimated by path analysis. Path analysis in twin studies usually include four latent variables: additive genes (A), common environment (C), unique environment (E) and genetic dominance (D). Monozygotic (MZ) twins have identical genes while dizygotic (DZ) twins share 50% of their genes. The genetic dominance in DZ twins correlate 0.25. Different models with various combinations of latent variables can be built to examine the genetic and environmental contributions to phenotypic variance.

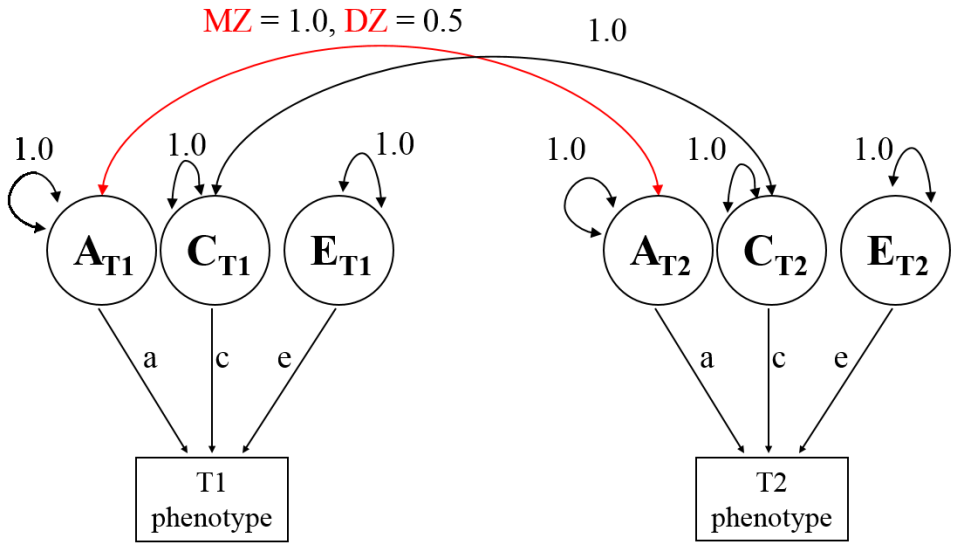
Take the ACE model as an example (I-Figure 6). If a-c-e are used to represent the path coefficient between corresponding latent variable (A-C-E) and a phenotype, then the explained variance of that phenotype by each latent variable will be: $Var_A = a^2$, $Var_C = c^2$ and $Var_E = e^2$. Using Cov to represent covariances from MZ and DZ twins, we will have the following equations:

(1) $Var_{phenotype} = a^2 + c^2 + e^2$

(2) $Cov_{MZ} = a^2 + c^2$

(3) $Cov_{DZ} = 0.5a^2 + c^2$

Equation (1) represents that the phenotypic variance is decomposed in genetic variance, shared and unshared environmental variance. Equation (2) represents that the covariance between MZ twins is explained by genetic and shared environmental variance. Equation (3) represents the covariance between DZ twins. Since DZ twins only share half of their genetic information, the genetic variance contributes only half compared to that in MZ twins.



I-Figure 6. Path diagram representing an ACE model, which hypothesises that phenotypic variance is only related to additive genetic (A), shared (C) and unshared (E) environmental factors. T1 and T2 represent the first-born and second-born twin, respectively. a, c and e represent the path coefficients between the corresponding latent variable (A, C and E) and the phenotype.

From those equations, we can calculate genetic variance (a^2). The heritability (h^2) is then calculated as $h^2 = a^2 / Var_{phenotype}$. As multiple models (e.g. AE, ADE model) can be built from the latent variables to test different hypotheses, the Akaike's information criterion (AIC) is usually used for model comparison regarding the same phenotype. The AIC evaluates the relative amount of information lost by a given model,[107] therefore, the model with the smallest AIC value will be the best model (among all candidate models) to explain phenotypic variance and calculate heritability. Using path analysis, various studies have been conducted to study the role of genetic variance in muscle strength and size phenotypes. For instance, Thomis et al.[108] reported that genetic factors accounted for 66–78% of the variance in arm strength, and Frederiksen et al.[60] found that hand grip strength had a heritability of 52%. The identification of a significant contribution of genetic factors

or heritability in a trait merits the further search for specific gene variants in measured genotype approaches.

1.6.2. Genetic association analysis

Genetic association analysis is often applied to identify genes/genetic variants that might be related to target traits or functions. To explore the association of a single genetic variant with a muscle mass/strength phenotype, comparisons of these muscular phenotypes between different groups of individuals with different genotypes for the genetic variant are usually conducted. For example, Roth et al.[78] used analysis of covariance (ANCOVA) to compare muscle mass and strength differences among *VDR* genotype groups after adjusting for factors such as age, physical activity and hormone levels, and found that GG carriers of the rs2228570 G/A polymorphism had less muscle mass and strength than the A allele carriers. Chi-square analysis showed that the distribution of the *VDR* GG homozygote was significantly higher in sarcopenic older men than in non-sarcopenic older men. Further logistic regression revealed that GG carriers had a 2.17-fold higher risk for sarcopenia than A allele carriers.[78]

In exploratory studies which aim to identify underlying genetic variants that are related to muscle mass and strength, association analyses are usually made on a large gene set. Few genome-wide association studies (GWASs) have been used in the identification of genetic variants that are related to muscle mass and strength in older adults. A GWAS is an observational study on a genome-wide set of genetic variants. Unlike the single genetic variant association analysis that has been discussed above, a GWAS is not driven by any candidate variants, and therefore, it can identify -hypothesis free- genetic variants that are closely related to a disease or a trait across the whole genome. Heckerman et al.[109] performed GWASs on physical performance of older adults (aged above 80 yrs) and identified two genetic variants (*ZNF295* rs928874 and *C2CD2* rs1788355) as being significantly related to the 4-meter gait speed. Tikkanen et al.[110] used a GWAS on hand grip strength of adults aged 40–69 yrs and found 101 loci associated with grip strength ($p < 5 \times 10^{-8}$). These loci were located in genes that code for molecules involved in neuro-developmental disorders or brain function. Two genetic variants (rs16892496 and rs7832552) from the *TRHR* gene were identified to be closely associated with LBM in a GWAS on one thousand US whites (aged above 50 yrs).[111] Individuals with unfavourable genotypes of rs16892496 and rs7832552 demonstrated 2.7 and 2.55

kg less LBM, respectively. Notably, because multiple comparisons are performed in a GWAS, a false discovery rate (FDR) (e.g. the Bonferroni correction and the Benjamini-Hochberg procedure) is required to control the expected rate of type I errors. Also due to the huge number of analysed genetic sites, a large sample size is needed to obtain enough statistical power. Based on genetic variants identified by a GWAS on muscle mass/strength, corresponding genes where those variants locate are determined. The biological functions and pathways that are related to those identified genes can be found through gene ontology (GO) enrichment and KEGG pathway analysis, which interpret possible biological domains (based on GO analysis) and pathways (based on KEGG analysis) that are related to particular gene sets.[112,113]

Repeated findings of genetic associations with replication of the specific beneficial allele are needed to strengthen gene-phenotype associations and minimize false positive findings. However, to prove causation between a specific sequence variation and a specific phenotype, more functional studies are needed. These can be done using different gene knockout methodologies in animal studies, informative cell lines, or other experimental setups. For example, gene knockout technology was used in *MSTN* (a gene encoding myostatin) knockout mice and showed greater increases in muscle mass (gastrocnemius and quadriceps) three months after myostatin depletion than in controls (wildtype mice), indicating a negative regulation on skeletal muscle mass associated with the activation of *MSTN*.[114] Contractile properties showed differences between individual muscle fibers of *ACTN3* RR and XX carriers, adding underlying knowledge in the observed association between ACTN3 R577X genotypes and power performance in athletes.[115]

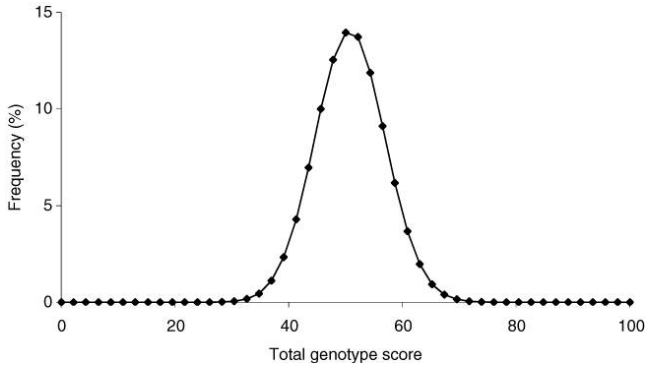
1.6.3. Genetic and methylation profile scores

Since multiple genetic variants are associated with ageing muscle,[66] the idea of building a genetic predisposing score is introduced to study a combined effect of multiple genotypes on a phenotype. In biomedical studies, a polygenic risk score (PRS) has been widely used due to its better predictive performance than significant SNPs identified by GWAS.[116] The PRS is calculated by summing weighted risk alleles, which are selected from GWAS results by certain p-value thresholds (e.g. $p < 1\times10^{-5}$).[117] The weight (also known as effect size) of a risk allele is determined by statistical methods such as Bayesian methods and penalized regressions.[116] In the field of sport science, the application of a genetic predisposing score was first used

by Williams and Folland[118] to estimate the distribution of people with endurance-favourable genotypes. In their study, 23 genetic polymorphisms were selected for their associations with endurance performance. A genetic score was given based on the genotype of each candidate gene: a homozygote that was positively associated with endurance phenotype was given a score of 2, a heterozygote got a score of 1 and the other homozygote was scored as 0. Given the 23 genetic scores GS_1, GS_2 … GS_{23}, a total genetic score (TGS) was further calculated by scaling the summed 23 genetic scores into a 0-100 range:

$$TGS = (100/46) \times (GS_1 + GS_2 + \ldots + GS_{23})$$

Using the frequency of each genotype reported in large sample-based studies, the researchers simulated the distribution of TGS based on 1 million hypothetical individuals and found that most individuals had similar endurance gene composition (I-Figure 7). Approximately 99% of individuals had a TGS of 37–65, indicating that most individuals differed by no more than seven genotype scores from the average.[118] Ruiz et al. applied the same TGS approach (based on seven endurance-related polymorphisms) to 46 world-class athletes in endurance activities (e.g. running, road cycling and rowing) and found that those athletes had a higher TGS than participants from a general population.[119,120] In **paper 4** of this thesis, the same approach was applied in the calculation of a TGS (GPS_{SNP}) by adding up genetic scores of seven muscle-related single nucleotide polymorphisms (SNPs) (*ACTN3* rs1815739, *ACE* rs4341, *CNTF* rs1800169, *FTO* rs9939609, *HIF1A* rs11549465, *MSTN* rs1805086 and *VDR* rs2228570).



I-Figure 7. Simulated distribution of endurance-related total genotype score (TGS, based on 23 endurance-related genes). The majority (99%) of individuals had a median TGS (37–65), very few individuals had a TGS at the two ends. Adapted from Williams and Folland, 2008.

The construction of a TGS allows researchers to compare combined genetic profiles among individuals, however, the TGS is applicable within a limited gene set, in which all polymorphisms are carefully selected under a criterion that their relationships with a phenotype have been reported in multiple studies. When analysing a target gene set with dozens or hundreds of variants, the TGS method is not ideal to be used because: (1) in a large gene set, some variants might have a weaker correlation with a phenotype than other variants; (2) the TGS will be affected by genetic variants that are closely correlated. For instance, if a genetic variant is positively associated with six other variants, then with a one-score addition of that variant, the TGS will be increased by seven units; (3) the TGS based on large sets of SNPs also leads to a leptokurtic distribution, i.e. the majority of subjects will have median TGSs while very few subjects will locate at both ends of the TGS distribution. Such a centralized distribution with a limited interindividual variation decreases the predictive power. Due to the reasons mentioned above, before the calculation of TGS, a stepwise regression is conducted to select genetic variants that are closely related to a phenotype and have comparatively weak correlations with each other. A genetic score calculated via such approach is known as a data-driven genetic predisposition score (GPS).[121] Bouchard reported that a data-driven GPS, which was calculated from 21 out of 39 polymorphisms, accounted for 49% of the variance in maximal O_2 uptake (VO_{2max}) trainability.[122] Thomaes et al. used backward regression to select muscle-related variants from 54 polymorphisms and reported that this data-driven GPS was positively related to knee muscle size and strength changes after a 3-month training in coronary artery disease (CAD) patients.[123] Charlier et al.[124] built a data-driven GPS based on 224 candidate genetic variants and demonstrated that this data-driven GPS explained up to 7% of the variance in muscle strength among 565 adults aged 19–73 yr. Such data-driven GPS approach (based on 170 candidate genetic variants) was used in **paper 1** and **2** of this thesis to study the genetic association with muscular phenotype changes in response to exercise and after the cessation of exercise.

A methylation profile score is a new approach that is recently introduced to explore the association of methylation levels at different CpG sites with disease and body composition.[125–128] Similar to the calculation of a total genetic score, a methylation score is calculated by summing methylation levels of particular CpG sites. A methylation level is a parameter which is used to represent the methylation status of a CpG site. It can be a beta (β) value or an M value. On a methylation detection

chip, each CpG site contains multiple probes to detect the methylation status of DNA fragments with this corresponding CpG site. The proportion of methylated probes is defined as a β value (β value = $\frac{No.of\ methylated\ probes}{Total\ No.of\ probes}$). A β value has a range of 0–1, and therefore, a total β value of 10 CpGs has a range of 0–10. An M value is a logit transformation of the ratio of methylated to unmethylated probes (M value = $\log\frac{No.of\ methylated\ probes}{No.of\ unmethylated\ probes}$). Due to the characteristics of a logarithm, an M value has an infinite range. A β value is an intuitive biological interpretation while an M value is more statistically valid for a differential methylation analysis.[129] Candidate CpGs that are used to calculate a methylation score can be predetermined or identified by various statistical approaches such as stepwise regression,[127] comparisons between different groups (with FDR correction)[126] and regressions with regularization (detailed description in section 1.4.4).[125] The application of methylation scores was reported to be helpful in improving the accuracy and sensitivity of diagnostic[126] and prognostic[127] prediction of prostate cancer. A BMI-related epigenetic score developed by Hamilton et al.[128] was also found to be associated with body mass, aerobic capacity, type 2 diabetes and cardiovascular disease. In **paper 4** of this thesis, one of the methylation scores (MS_{SNP}) was calculated based on the average methylation level of CpGs that located within the predetermined seven muscle-related genes as previously mentioned.

Moreover, recent studies have combined genetic and methylation profiles to explore hereditary and environmental associations with physical conditions such as BMI and heart disease risk. Shah et al.[130] found that regression models with only BMI-derived genetic or methylation scores explained less than 11% of the inter-individual variance in BMI, while a model combining both scores improved the explained variance to 13–18%. Dogan et al. also reported that a model with integrated genetic and methylation scores outperformed (with 13% more accuracy) conventional risk factors in predicting coronary heart disease.[131] By far, **no findings on combined genetic and methylation profile scores in muscle mass and strength of older adults have been reported**.

1.6.4. Genetic variant and CpG site selection using Elastic net and Lasso regularization

In the calculation of TGS and data-driven GPS, selected genetic variants are equally weighted. Yet, genetic variants might contribute differently to a phenotype. Moreover,

both TGS and data-driven GPS approaches have limited statistical power, which means that the sample size of genotyped individuals needs to be much larger than the number of candidate genetic variants. Therefore, new approaches (e.g. lasso regularization and elastic net regularization) are applied for variable selection and weight estimation when the candidate genetic/methylation variables outnumber the sample size.

To restrict overfitting and ensure a good predictive power, a regularization on coefficients is often used in a regression model. A regression model with a L^1 regularization (the Manhattan norm) is called lasso regression. A regression model with a L^2 regularization (the Euclidean norm) is called ridge regression. Take the lasso linear regression as an example. A conventional linear regression model estimates coefficients (β) of independent variables (x) by minimizing the residual sum of squares (RSS), which can be presented as follows:

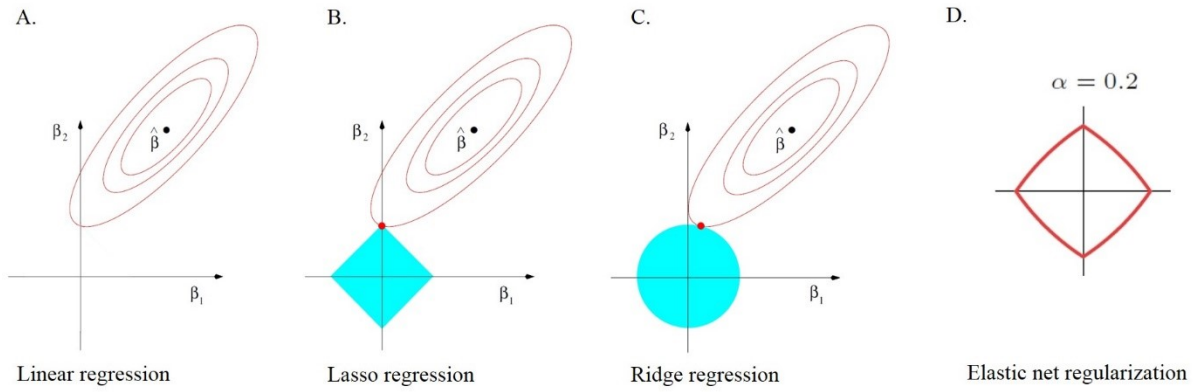
f(1) = $\min_{\beta_0,\beta}\{\frac{1}{N}\sum_{i=1}^{N}(y^i - (\beta_0 + X^i\beta))^2\}$, where *i* represents the *i*th sample.

Similarly, a lasso linear regression estimates coefficients of independent variables by minimizing a function combining the RSS and the L^1 regularization on coefficients, which can be represented as follows:

f(2) = $\min_{\beta_0,\beta}\{\frac{1}{N}\sum_{i=1}^{N}\left(y^i - (\beta_0 + X^i\beta)\right)^2 + \lambda\sum|\beta|\}$, where *i* represents the *i*th sample and λ is the shrinkage parameter. The larger the value of λ is, the greater extent of shrinkage the coefficients will have.

Therefore, the lasso linear regression aims to minimize RSS by selecting coefficients from a constraint region: $\sum|\beta| \leq \frac{1}{\lambda}$. To better demonstrate the regularization process on coefficients, let's suppose there are two independent variables (x_1 and x_2) with corresponding coefficients (β_1 and β_2) in a linear regression model. The estimation of optimal coefficients can be illustrated as in I-Figure 8B. The x and y axis represent β_1 and β_2, respectively. $\hat{\beta}$ is the optimal point calculated from function f(1) (I-Figure 8A), where the least RSS is reached in a conventional linear regression. The red ellipses are the contours of the RSS function. β_1 and β_2 values on the same ellipse yield the same RSS. The larger an ellipse is, the higher value of RSS it represents. In a L^1 regularization, the coefficients (β_1 and β_2) are constrained in a region highlighted as a light blue square. In order to minimize the RSS, the first intersection of an ellipse and the constraint region will

give the optimal coefficients. Considering the characteristics of a square-shape constraint region in lasso regression, the intersection (represented as a red dot) is often on one of the axis, meaning one of the coefficients will be zero. Through this process, variables that are strongly related to a phenotype (with a large coefficient that is resistant to shrinkage) will be selected and the corresponding coefficient will be the weight of that variable.



I-Figure 8. Regularization for variable selection. (A) Conventional linear regression. Red ellipses represent the contours of the residual sum of squares (RSS). (B) Coefficient selection by lasso regression. (C) Coefficient selection by ridge regression. (D) Constraint region in elastic net regularization (α = 0.2). Adapted from Elements of statistics learning, Second Edition, 2009.

The application of lasso regression for variable selection can be found in several methylation studies. Reese et al. used a genome-wide lasso logistic regression and identified 28 CpGs to build a predictive model for maternal smoking during pregnancy with an accuracy of 91% and a specificity of 97%.[132] Based on the methylation of five CpG sites selected from 450K candidate CpGs by lasso regression, Zhao et al.[125] built a predictive model for clear cell renal cell carcinoma (CCRCC) prognosis and the model showed reliable predictions across several cohorts.

Similar to the lasso regression, the L^2 regularization (ridge regression) aims to minimize the following function:

$$f(3) = \min_{\beta_0,\beta}\{\frac{1}{N}\sum_{i=1}^{N}\left(y^i - \left(\beta_0 + X^i\beta\right)\right)^2 + \lambda\sum\beta^2\}$$

Since the constraint region is a circle, its intersection with an RSS ellipse is less likely to be on an axis (I-Figure 8C). Therefore, most variables still remain in the

model. Due to the mild regularization on coefficients, the ridge regression is rarely used in variable selection.

The elastic net regularization combines both lasso and ridge regression by giving different weights to the L^1 and L^2 regularization. It aims to minimize the following function:

f(4) = $\min_{\beta_0,\beta}\{\frac{1}{N}\sum_{i=1}^{N}\left(y^i-\left(\beta_0+X^i\beta\right)\right)^2+(1-\alpha)\sum\beta^2+\alpha\sum|\beta|\}$, where α is in a range 0–1.

The elastic net regularization has a constraint strength between the lasso and ridge regression, and is also used in genetic variant selection (I-Figure 8D). Cho et al.[133] applied the elastic net approach in a genome-wide association analysis and identified 129 genetic variants that were associated with adult height in a Korean population. They found that with a one unit increase of the genetic score built by these genetic variants, the average height was increased by 0.47cm.[133]

In the current thesis, the lasso method was combined with a logistic regression procedure in **paper 4** to select sarcopenia-driven CpGs. In the lasso logistic regression model, the sarcopenia status (sarcopenia was coded as 1 and non-sarcopenia was coded as 0) was the dependent variable, and the methylation level of each measured CpG was the independent variable. The methylation levels and weights (obtained from the lasso logistic regression) of selected CpGs were subsequently used in the calculation of the sarcopenia-driven methylation score (MS_{SAR}).

2. Outline of studies in this thesis

As previously discussed, inter-individual variability in muscle mass and strength exists among older adults during the ageing process. Although genetic and DNA methylation profiles are probably related to muscle characteristics at an older age, studies focusing on explaining this inter-individual variability in muscle mass and strength by genetic architecture and DNA methylation are lacking. Therefore, this thesis performs cross-sectional analyses in older female adults to explore the role of DNA sequence variation and DNA methylation in ageing muscle. Moreover, since inter-individual differences are also found in muscle changes after exercise intervention and after the cessation of exercise, this thesis studies the role of muscle-related genetic variants in muscle mass and strength changes during

exercise and detraining among older adults who received a one-year exercise intervention and experienced a one-year cessation of training.

The current thesis consists of three chapters. **Chapter 1** discusses the association between genetic variants and exercise-related muscle changes in older people. This chapter includes two studies, in which the genetic and muscular data are collected from a one-year intervention and follow-up study within the framework of the first Policy Research Center *Sport, Beweging en Gezondheid* at the KU Leuven ("Thema 3.6" - "Effect van een oefenprogramma op gezondheid- en fitnessgerelateerde parameters in een groep van ouderen."). **Chapter 2** identifies sarcopenia-related DNA methylation differences based on methylation profiles of blood cells. **Chapter 3** studies the association between genetic architecture, DNA methylation and ageing muscle by analysing linear models with integrated genetic and methylation profile scores. **Chapter 2** and **3** are based on data collected from the same study on older women through the Manchester Metropolitan University (MMU) project "Genetics of sarcopenia".

2.1. Chapter 1: Genetic variants and exercise-related muscle changes in older people

2.1.1. Paper 1: Genetic predisposition score predicts the increases of knee strength and muscle mass after one-year exercise in healthy elderly

Inter-individual variability is observed not only in muscle mass and strength among individuals, but also in their adaptations to exercise. Such variability might be partly related to genetic architecture of an individual since many studies have reported a significant hereditary factor in muscle mass, strength, and physical performance. Therefore, the aim of this study is to explore the relationship between genetic profile scores and baseline muscular phenotypes (muscle mass and strength) as well as muscle adaptations after a one-year supervised exercise intervention in 200 Flemish Caucasians aged 60–83 yrs. Using data-driven GPSs derived from 170 candidate genetic variants (as described in section 1.4.3), this study analyses the percentage of variance in baseline muscular parameters and exercise-induced muscle adaptations explained by a GPS.

Hypotheses of paper 1

- Data-driven GPS has a significantly predictive value in baseline muscle mass and strength in older adults. An individual with a higher GPS will have a larger muscle mass and strength.
- Muscle mass and strength will increase after training, however, substantial inter-individual variability in these muscular response phenotypes exists.
- Data-driven GPS is positively associated with muscle adaptations to exercise. An individual with a higher GPS will have a larger muscular improvement after structured training.

2.1.2. Paper 2: The genetic association with muscular changes in an older population after one-year of detraining: a follow-up study after one-year cessation of structured training

Muscle mass and strength usually decline after the cessation of structured training. Notably, the decrease in muscle mass and strength is not consistent among individuals and the genetic basis of such inter-individual variability has never been studied. Therefore, this study analyses the possible connection between an individual's genetic architecture and muscular changes after one-year cessation of structured training among older adults.

Hypotheses of paper 2

- Muscle mass and strength will decrease after a one-year cessation of training with inter-individual differences.
- Data-driven GPS is negatively related to the decline in muscle mass and strength after the cessation of training. An individual with a higher GPS will have smaller loss in muscle mass and strength.

2.2. Chapter 2: Sarcopenia-related DNA methylation differences

Paper 3: Differentially methylated gene patterns between age-matched sarcopenic and non-sarcopenic women

Muscle mass and strength are determined by both genetic and environmental factors. As previously introduced, DNA methylation is one of the main mechanisms for environmental factors to regulate gene expression. Therefore, this study aims at discovering DNA methylation patterns that are related to inter-individual variability in muscular fitness (muscle mass and strength). To better identify those methylation

patterns, this study compares blood DNA methylation profiles of 788K CpG sites between sarcopenic and non-sarcopenic older women that are characterised by significant differences in muscle mass and strength. Although sarcopenia is a muscle disorder, it is also influenced by the circulatory and endocrine system as multiple cytokines and hormones influence muscle homeostasis.[134–136] The methylation patterns of our participants are analysed based on whole blood DNA samples, an easily accessible tissue, compared to DNA from muscle biopsies. Due to a limited sample size, this study has an explorative nature to detect a set of sarcopenia-related methylation differences in blood samples and to evaluate the feasibility of using blood DNA methylation to identify systemic, circulatory and muscle-specific differentially methylated markers of ageing muscle.

Research questions of paper 3

- To identify sarcopenia-related genes and CpGs that demonstrate differentially methylated patterns between sarcopenic and non-sarcopenic older women.
- To explore possible sarcopenia-related biological functions and pathways based on sarcopenia-related genes.

2.3. Chapter 3: Role of genetic and methylation profiles in ageing muscle phenotypes

Paper 4: Associations of combined genetic and methylation profile scores with muscle size and strength: a pilot study in older women

This is an explanatory study of both genetic and DNA methylation profile scores with muscle size and strength in older women. Different linear models are built based on genetic and methylation profile scores with muscular phenotypes (muscle morphology and strength) as dependent variables. A genetic score (GPS_{SNP}, as described in section 1.4.3) is calculated based on seven genetic variants (*ACTN3* rs1815739, *ACE* rs4341, *CNTF* rs1800169, *FTO* rs9939609, *HIF1A* rs11549465, *MSTN* rs1805086 and *VDR* rs2228570) that are previously reported in at least three papers as being closely related to skeletal muscle with a consistent direction of the favourable genotype. A methylation score is calculated based on the methylation levels of (1) CpGs selected by a sarcopenia-driven lasso logistic regression (MS_{SAR}, as described in section 1.4.4) or (2) CpGs in genes where the seven genetic variants locate (MS_{SNP}, as described in section 1.4.3).

Hypotheses of paper 4

- Both genetic and methylation profile scores have a significant predictive value to explain muscle size and strength variability in older women.
- A model with both genetic *and* methylation profile scores explains more inter-individual variability in muscle morphology and strength than a model with only a genetic *or* methylation profile score.

3. Study design and methodology

3.1. Participants

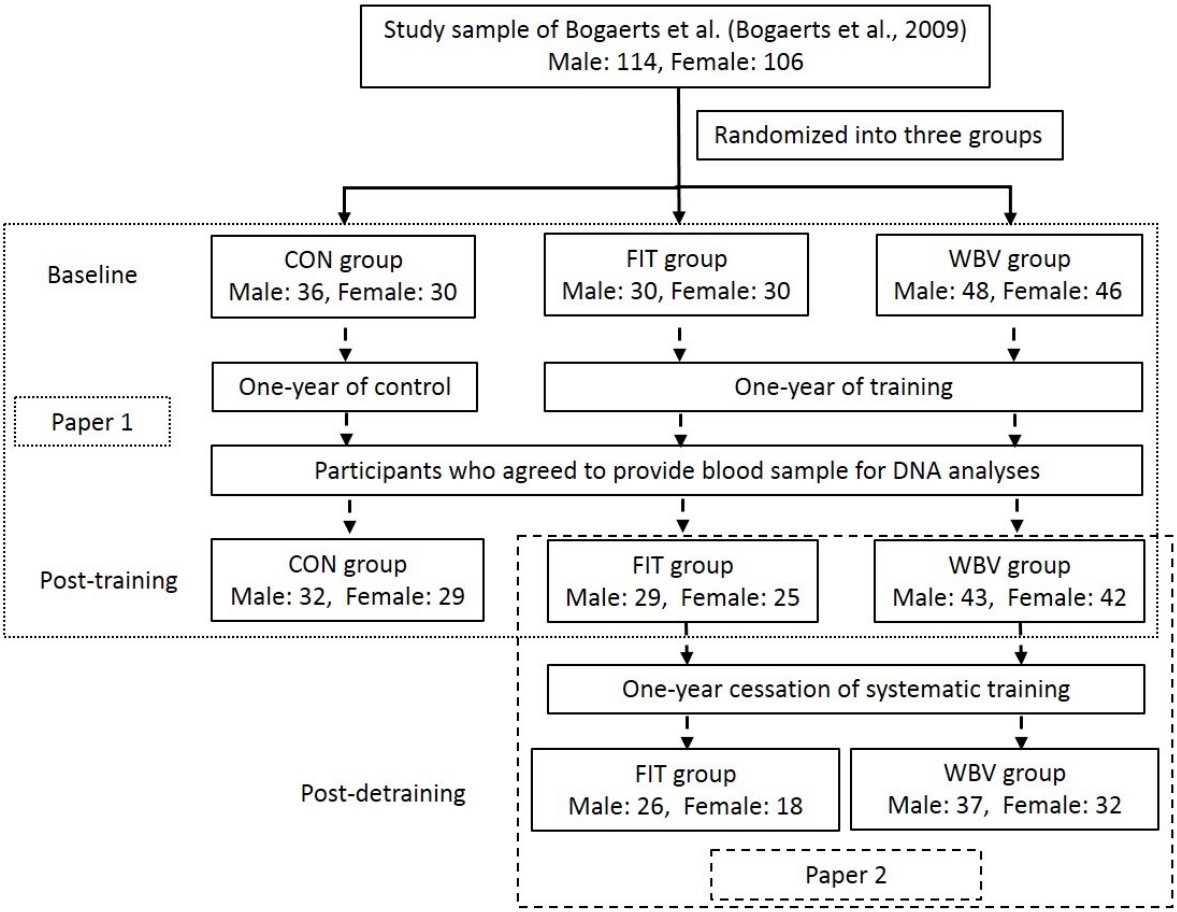
3.1.1. Paper 1 and 2: First Policy Research Center 'Sport, Beweging en Gezondheid' - KULeuven project "Theme 3.6"

Participant data used in the first and second study were collected from the KU Leuven project "Theme 3.6: Effect van een oefenprogramma op gezondheid- en fitnessgerelateerde parameters in een groep van ouderen." (2004-2012). In this project, older people aged between 60 to 83 years were recruited from local communities around the city of Leuven. All participants went through a series of medical examinations. Participants with skeletal, neuromuscular or cardiovascular disorders (e.g. rheumatoid arthritis and Alzheimer's disease) that may impede strength training and muscle strength tests, or with training experience in the past two years were excluded. The selected participants were randomly assigned into a fitness group (with combined resistance and aerobic training, FIT), a whole-body vibration (WBV) group and a control (CON) group. Participants in the two exercise groups (FIT and WBV) conducted a one-year structured training (training protocols in P1-Supplementary Table 1). 200 participants (104 men, 96 women) provided a blood sample for DNA analyses and their data were analysed in **paper 1** (I-Figure 9). Among the 200 participants, 54 of them performed the fitness program (FIT group), 85 of them were in the WBV group and the rest were control subjects (CON group) (I-Table 1). One year after the training program, participants who had provided blood samples in the exercise groups were contacted for a follow-up test. The data collected in the follow-up test were analysed in **paper 2** (I-Table 1, I-Figure 9). This project was approved by the University's Human Ethics Committee in accordance with the Declaration of Helsinki. Informed consent was given by each subject. Data analysis on primary training outcomes have been published within the doctoral thesis of Bogaerts A.[137,138]

I-Table 1. Descriptive characteristics of participants in **paper 1** and **2**.

Paper			Paper 1 (one-year of training)			Paper 2 (one-year after the cessation of training)	
Group			CON	FIT	WBV	FIT	WBV
Number	F		29	25	42	18	32
	M		32	29	43	26	37
Age (yr)	F		68 ± 5	66 ± 4	67 ± 5	66 ± 4	66 ± 5
	M		69 ± 6	67 ± 4	68 ± 5	68 ± 4	67 ± 4
Height (cm)	F		161.3 ± 7.0	160.0 ± 7.9	161.2 ± 5.7	158.5 ± 7.7	160.7 ± 6.1
	M		173.0 ± 5.4	174.3 ± 6.1	173.1 ± 6.5	175.1 ± 5.9	173.4 ± 6.5
Body mass	F	Pre-training	69.2 ± 9.9	67.9 ± 9.2	69.3 ± 8.9	67.1 ± 9.9	67.9 ± 7.2
(kg)		Post-training	68.7 ± 10.0	66.6 ± 9.4	68.7 ± 9.3	66.1 ± 10.2	67.3 ± 7.7
		Detraining	-	-	-	66.2 ± 9.9	68.2 ± 7.7
	M	Pre-training	80.9 ± 8.5	83.2 ± 9.3	81.0 ± 13.1	82.7 ± 9.1	81.4 ± 13.1
		Post-training	79.7 ± 8.7	82.0 ± 9.6	78.9 ± 11.6	81.2 ± 8.9	79.2 ± 11.2
		Detraining	-	-	-	82.4 ± 9.0	80.9 ± 12.7

CON: control group; FIT: fitness group; WBV: whole-body vibration group



I-Figure 9. Flowchart of participants in paper 1 and 2. Paper 1 includes 200 participants (104 men, 96 women) and paper 2 includes 113 participants (63 men, 50 women).

3.1.2. Paper 3 and 4: MMU project "Genetics of sarcopenia"

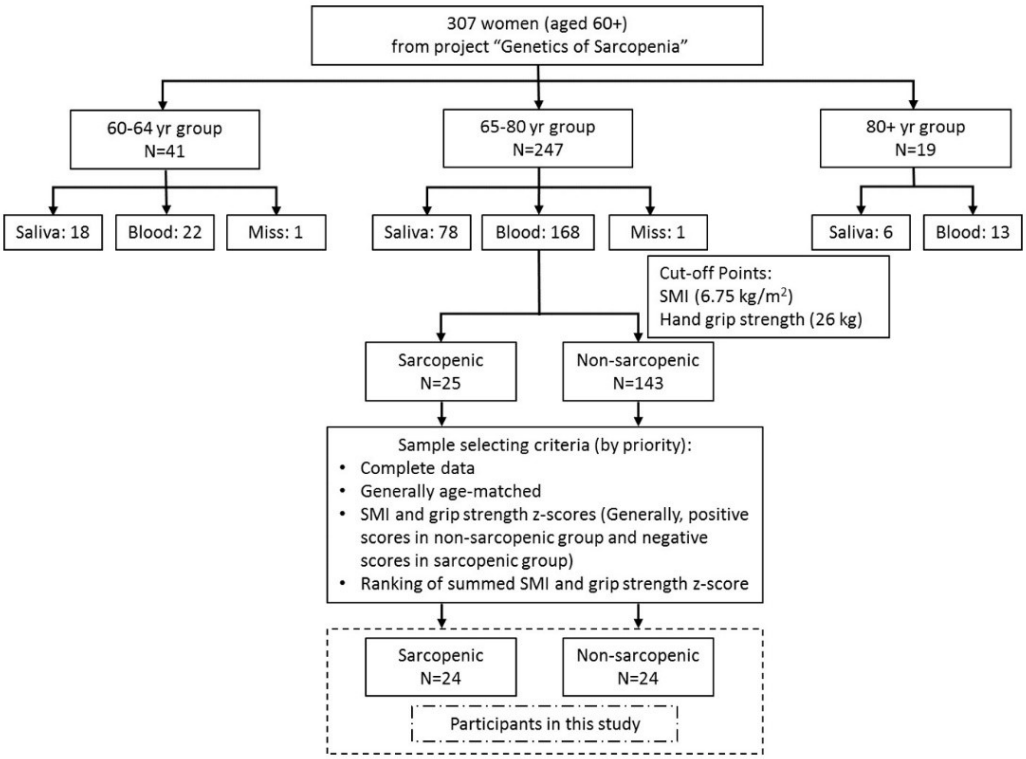
Participants in **paper 3** and **4** were selected from an initial set of 247 older, independently living, Caucasian women (aged 65–80 yr) recruited from the local area of Crewe (Move-age mobility period: 2016–2017, Manchester Metropolitan University, Crewe, UK). 168 participants provided 5 mL venous blood samples and those participants were subsequently categorised into sarcopenic (n = 25) and non-sarcopenic (n = 138) groups using cut-off points of skeletal muscle index (SMI, calculated using skeletal muscle mass divided by height squared) at 6.75 kg/m^2 [16] and hand grip strength (HGS) at 26 kg (the lower quintile of HGS in the total participant group). Through a process of further selection including age matching,[98] completeness of data, rankings of SMI and hand grip strength z score, and summed z score (I-Figure 10), 24 participants (age of sarcopenic group 72.5 ± 4.2 yr, non-

sarcopenic group 70.5 ± 3.3 yr) from each group were selected for DNA methylation analysis. In the sarcopenic group, 21 participants with negative z scores in SMI and HGS were selected first, with an additional three selected via an ascending sequence of summed z scores. Selection in the non-sarcopenic group was done in an opposite direction: 23 participants with positive z scores in SMI and HGS were selected first, with an additional participant with the highest summed z score selected from the remainder (I-Figure 11). Therefore, the sample size in **paper 3** and **4** is 48 (24 sarcopenic, 24 non-sarcopenic, I-Table 2).

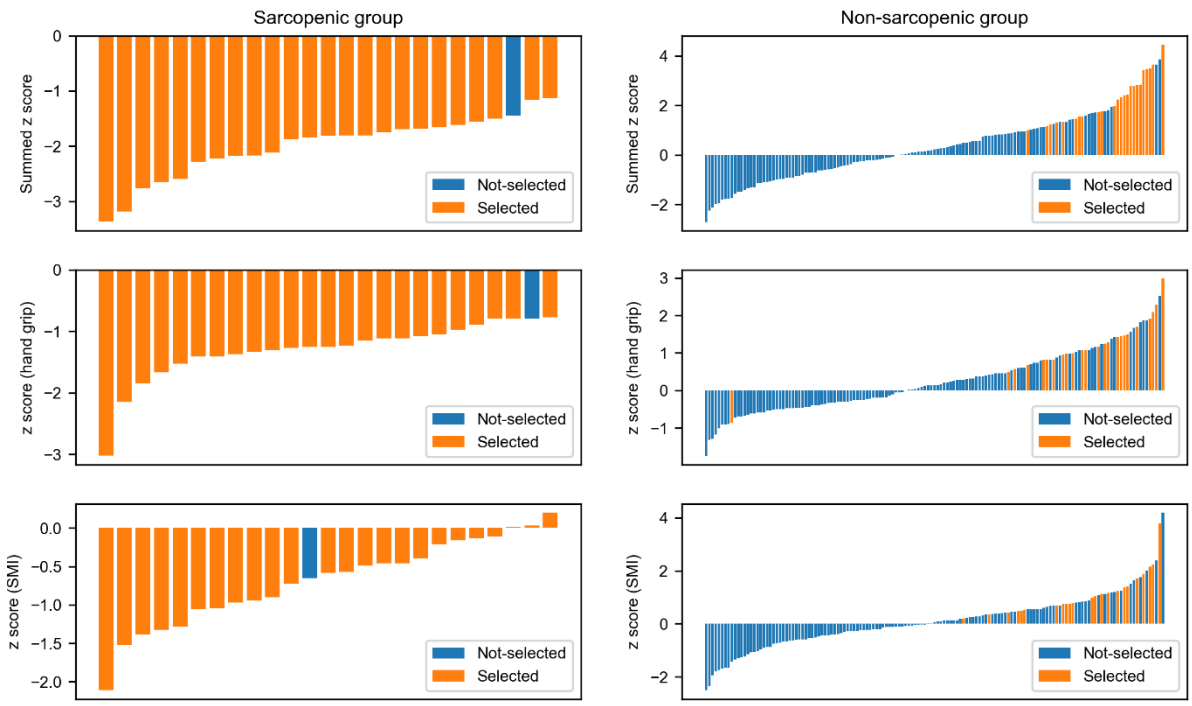
I-Table 2. Descriptive characteristics of participants in **paper 3** and **4**.

Group	Non-sarcopenic	Sarcopenic	Total group
Number	24	24	247
Age (yr)	70 ± 3	73 ± 4	71 ± 4
Height (m)	1.60 ± 0.05	1.56 ± 0.11	1.59 ± 0.06
Body mass (kg)	71.7 ± 12.8*	61.5 ± 9.4	66.6 ± 12.3
Grip strength (kg)	36.0 ± 3.7*	23.2 ± 2.5	29.6 ± 7.1*
SMI (kg/m^2)	7.45 ± 0.67*	6.00 ± 0.47	6.72 ± 0.93*

*: significant difference compared with the sarcopenic group (p < 0.01)



I-Figure 10. Flowchart of participant selection for paper 3 and 4. 48 participants are selected from an initial sample group of 247 with SMI and grip strength as criteria.



I-Figure 11. Distribution of z scores on SMI, handgrip strength and summed z score for participant selection in the MMU project. Participants (n=21) with negative z scores in both SMI and hand grip strength are firstly selected from the sarcopenic group. An additional three are selected via an ascending sequence of summed z scores. Participants (n=23) with positive z scores in both SMI and hand grip strength are firstly selected from the non-sarcopenic group. An additional participant with the highest summed z score is selected from the remainder.

3.2. Methodology

3.2.1. Methodology in the KULeuven project “Theme 3.6” – paper 1 and 2.

3.2.1.1. Genotyping

A 4.5 ml fasting blood sample of each participant was drawn from an antecubital vein in an EDTA-coated tube. Genomic DNA was extracted using the chemagic Magnetic Separation Module I (chamagic MSM I, PerkinElmer Inc., Waltham, MA, USA). Genotyping was done with the Illumina GoldenGate platform (Illumina Inc., San Diego, CA, USA) at the Genomics Core Facility (UZ/KU Leuven). 224 muscle-related genetic variants were genotyped (P1-Supplementary Table 2, detailed description in section 3.2.1.2). Through blood testing, 12 SNPs were not successfully detected and 3 SNPs presented the same genotypes among all subjects (probably due to the high frequency of dominant alleles, P1-Supplementary Table 2). Those 15 SNPs were ruled out from the 224-SNP pool. Results of linkage

disequilibrium tests showed that 58 SNPs were highly linked as 19 subgroups and one representative SNP was selected from each of these subgroups. A total number of 170 SNPs were kept for further analyses.

3.2.1.2. Candidate genetic variants selection

The selection of genetic variants was based on published articles (up to August 2014) and expression quantitative trait loci (eQTL) analysis. Muscle-related SNPs and genes were first searched in the PubMed and the Medline databases using multiple keywords (e.g. polymorphism, genotype, muscle mass and strength). SNPs and genes that were significantly associated with at least one muscular phenotype were included. Genes that were previously selected by literature search were further put into the GeneVar platform (Trust Sanger Institute, https://www.sanger.ac.uk/science/tools/genevar-gene-expression-variation-archive) to search for eQTLs of corresponding genes. An eQTL is a locus that explains a fraction of variance in expression levels of mRNAs,[139] therefore, it might have a more functional role in muscle mass and strength than other genetic variants. Since eQTLs in muscle tissue were not reported in the GeneVar platform, candidate eQTLs were selected if they showed significance in the GeneVar cis-eQTL analysis in at least two different tissues with a p-value < 0.0001.

In the genetic analysis on muscular phenotypes, a data-driven GPS approach was used. This part of methodology has been explained in section 1.4.3. Briefly, stepwise regression was used (entry significance: 0.1, stay significance: 0.05) with 170 SNPs as independent variables and a specific muscular phenotype as the dependent variable. SNPs that were selected were regarded as data-driven SNPs of the corresponding muscular phenotype and a data-driven GPS was calculated by summing the genetic score of each data-driven SNP.

3.2.1.3. Muscular phenotype measurement

Whole-body skeletal muscle mass (SMM) was calculated through bioelectrical impedance analysis (BIA). Resistance of BIA was measured by Bodystat 1500MDD (Bodystat Ltd, Douglas, UK) before and after the one-year intervention. Before the test, participants were asked to lie down in a supine position for one minute. During the measurement, two electrodes were placed on the right hand and right foot as instructed in the manual. SMM was calculated for further analyses, using the following regression equation that has been assessed for validity in elderly participants[140]: SMM (kg) = ($Ht^2/R \times 0.401$) + (sex × 3.825) + [age × (-0.071)] +

5.102 where Ht stands for height in centimeters; R stands for BIA resistance in ohms; in sex, men = 1 and women = 0; age is in years.

Biodex Medical System 3 dynamometer (Biodex Company, New York, USA) was used for the measurement of isometric knee extensor strength. This measurement was done by the same operator before and after the intervention. Before testing, participants were asked to complete a 5-minute warm up on a free-loaded ergometer. Two practice trials were performed to allow for a better understanding of the measuring process. Maximal isometric knee extension strength was evaluated at knee flexion angles of 60° with 0° representing full extension. Peak velocity of knee extension movement was measured by isotonic test with a load of 20% of the peak isometric strength obtained at the knee flexion angle of 90°. Participants were asked to extend their legs with full effort until they achieved the knee flexion angle of 20°. In isokinetic tests, participants performed isokinetic knee extension and flexion movements at two different speeds. The first measurement required participants to complete four repetitions at a low velocity of 60°/s. The second measurement consisted of six repetitions at a higher velocity of 240°/s. Peak torque of knee extensors at 60°/s and at 240°/s were recorded as isokinetic knee strength.

3.2.2. Methodology in the MMU project “Genetics of sarcopenia”

3.2.2.1. Genotyping

A 5 mL venous blood sample was collected from each participant and stored in an EDTA-coated tube at -20 Celsius for DNA extraction. DNA samples were extracted using QIAcube® and QIAamp® DNA Blood Mini Kit (Qiagen, Crawley, UK) according to the manufacturer’s instructions. Briefly, the spin column and 1.5 ml centrifuge tube were first put into the rotor adaptor. The sample tube with 200 µl blood sample (defrost under room temperature) was put into the shaker rack. AW1 (ethanol diluted) was put into the reagent bottle rack and the protease was put into the tip rack. Finally, the DNA extraction was performed under the “100 µl blood sample test” model. The extracted DNA samples were stored at -20 Celsius for genotyping and DNA methylation analysis.

Seven SNPs (described in section 2.3) were selected for genotyping. These SNPs have been reported in at least three papers as being closely related to muscle strength or mass with a consistent direction of favourable alleles (P4-Supplementary Table 1). Duplicate genotyping was firstly made using a 192.24 Dynamic Array® IFC (Fluidigm Corp., South San Francisco, CA, US) and TaqMan SNP genotyping

assays (Applied Biosystems, Paisley, UK) following the manufacturer's instructions. Briefly, a genotyping mix (4 µL) consisted of 2 µL assay loading reagent [2x] (Fluidigm), 1 µL SNP genotyping Assay Mix [40X] (Applied Biosystems), 0.2 µL ROX [50X] (Invitrogen, Carlsbad, CA, US) and 0.8 µL DNA-free water (Qiagen). A sample mix (4 µL) contained 1.6 µL DNA samples, 2.0 µL GTXpress master mix [2X] (Applied Biosystems, PN 4401892), 0.2 µL Fast GT Sample Loading Reagent [20X] (Fluidigm, PN 100–3065), and 0.2 µL DNA-free water. All reaction mixes (7.75 µL, consisting of 3.75 µL genotyping mix and 4 µL sample mix) were loaded onto the Dynamic Array IFC following the manufacturer's instructions. The array was subsequently placed into a thermal cycler (FC1 Fluidigm, PN 100-1279 D1) and the GT 192.24 Fast v1.pcl protocol was performed. The thermal cycling protocol included an amplification at 95 °C for 120 s followed by 45 cycles of denaturation for 2 s at 95 °C and extension for 20 s at 60 °C. Reporter dyes VIC and FAM were used for genotyping based on fluorescence detection.

About 1% of SNP-sample data points showed unsuccessful detection or inconsistent genotype results using the Fluidigm system. These SNP samples were reassessed in duplicates using a StepOnePlus Real-Time PCR system with TaqMan SNP genotyping assays and analysed using StepOnePlus analysis software (Applied Biosystems, version 2.3). The StepOnePlus reaction mix (10 µL) included 0.2 µL DNA sample, 5 µL GTXpress master mix, 4.3 µL nuclease-free water and 0.5 µL TaqMan SNP genotyping assay [20X]. Each reaction mix was amplified for 20 s at 95 °C, followed by 50 cycles of denaturation for 3 s at 95 °C and extension for 20 s at 60 °C. Genotypes were identified based on fluorescence detection of reporter dyes (VIC and FAM).

3.2.2.2. DNA methylation analysis

DNA methylation (blood-based) was measured using Illumina® Infinium MethylationEPIC BeadChip arrays (Illumina Inc., San Diego, CA, US) at the Genomics Core facility (UZ/KU Leuven). DNA samples were quantified by Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, US) and diluted into 10-25 ng/µl by distilled water. Bisulfite conversion of DNA was undertaken by EZ-96 DNA Methylation-Direct Kit (Zymo Research, Irvine, CA, US). 130 µl CT conversion reagent solution was mixed with 20 µl sample in the conversion plate. The plate was later put into a thermal cycler and was treated at 98 °C for 8 minutes, 64°C for 3.5 hours and stayed at 4 °C overnight. After the conversion and amplification, the

mixed solution was transported to Infinium MethylationEPIC Beadchip (Illumina, San Diego, CA, United States) and was stained by STM, ATM and XC3 for three times. The stained Beadchip was scanned Genome Studio Methylation Module v1.8 (Illumina, San Diego, CA, United States) which presented the green/red fluorescence intensity as methylation files.

Those methylation files were later analysed by R 'Minfi' package,[141] background signals were corrected by normal-exponential out-of-band (Noob) method, and methylation values (β values) were normalized for blood cell composition by R 'FlowSorted.Blood.EPIC' package.[142] Probes were dropped under one of the three conditions: 1) probes with non-significant background signal levels (p > 0.01) at methylated and unmethylated channels; 2) probes that contain either single nucleotide polymorphisms at the CpG interrogation or at the single nucleotide extension as suggested in the "Minfi" package (reference array: "IlluminaHumanMethylationEPIC", annotated by ilm10b4.hg19); 3) cross-active probes that were reported in the first supplementary table of Pidsley's study.[143] A final total of 788,074 probes were kept for further methylation analyses using Partek Genomics Suite V.7.0 (Partek Inc., St. Louis, MO, US), in which CpG probes were annotated based on "HumanMethylation850" reference, "MethylationEPIC_v-1-0_B4" annotation file, "Homo sapiens" species and hg19 genome build. Notably, the differential methylation analysis between the sarcopenic and non-sarcopenic groups (**paper 3**) was based on the M value (as described in section 1.4.3), and CpGs with significantly different methylation levels between the sarcopenic and non-sarcopenic groups were defined as differentially methylated CpGs (dmCpGs). Genes where dmCpGs located were further analysed by GO enrichment and KEGG pathway analysis using Partek to determine sarcopenia-related biological functions and pathways.

A sarcopenia-driven methylation profile score (MS_{SAR}, as described in section 1.4.4) in **paper 4** was calculated from CpGs selected by a sarcopenia-driven lasso logistic regression. The mechanism of the lasso regression has been described in section 1.4.4. In **paper 4**, a six-fold cross validation (with the log loss score, the accuracy score and the F1 score as metrics) was used for the shrinkage parameter (λ) tuning (P4-Supplementary Table 2A). Sarcopenia-driven CpGs were selected as those with non-zero coefficients in the lasso regression with an optimal shrinkage parameter (λ = 65.1318, P4-Supplementary Table 2A). The sarcopenia-driven methylation score was calculated as a weighted sum of the selected CpG

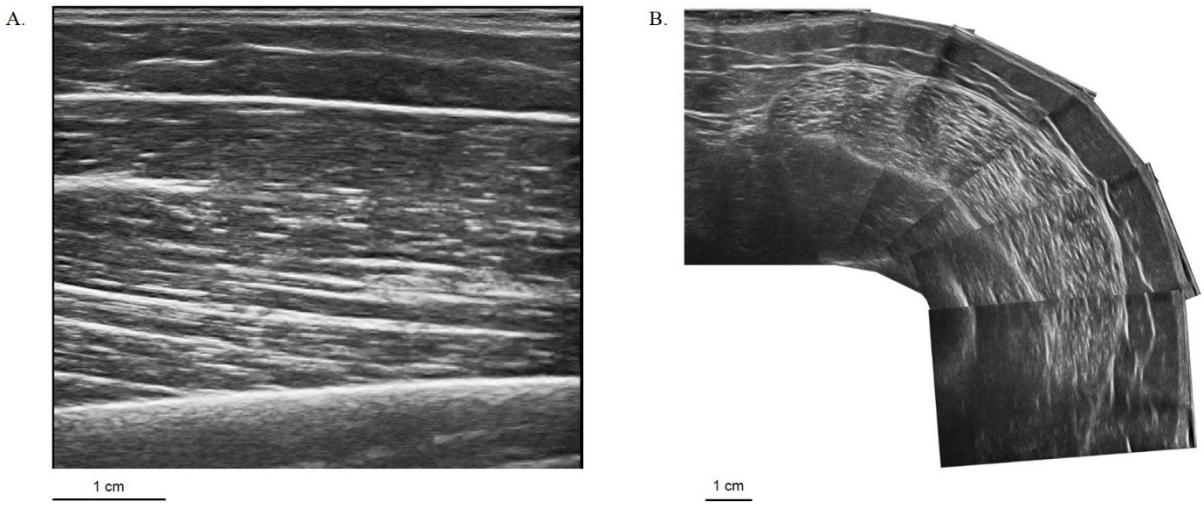
methylation levels (the weight for each CpG site was the coefficient from the lasso regression, codes in P4-Supplementary File 1). In **paper 4**, another methylation profile score (MS_{SNP}, as described in section 1.4.3) was calculated based on the average methylation level of CpGs that located within the predetermined seven muscle-related genes (described in section 2.3). Additionally, to explore the genetic association of each muscle-related gene with muscle size and strength, a gene-wise combined genetic and methylation profile score was created as the ratio of a SNP score to the average methylation level in promoters of the corresponding gene.

3.2.2.3. Muscular phenotype measurement

Whole-body skeletal muscle mass was estimated using the same equipment and equation described in section 3.2.1.3. In the MMU project, hand grip strength was also measured. Participants were asked to stand straight and to keep their testing arms straight out during the measurement. Verbal encouragement was given and three attempts were made on both hands. The highest value was kept for further analysis.

Biceps brachii thickness (THK_{BB}) and vastus lateralis (VL) anatomical cross-sectional area ($ACSA_{VL}$) were measured by B-mode ultrasonography (MyLab®Twice Esaote, Genoa, Italy). During the measurement of THK_{BB}, participants were sitting with their elbows extended and relaxed. Sagittal plane scans were taken and the muscle thickness was measured at three sites: 60% of the length from the acromion process of the scapula to the lateral epicondyle of the humerus[144], and the upper and lower site 1 cm away from the 60%-length site. The muscle thickness was measured using an image processing program (ImageJ, NIH) by the same investigator (I-Figure 12A, intraclass correlation coefficient [ICC] = 0.98, the ICC was an interrater reliability based on a single scan that was assessed twice among six participants). Measurement of $ACSA_{VL}$ was performed in a seating position. Axial plane scans were taken at 50% muscle length of the VL and recorded in real time, with the ultrasound probe passing over echo-absorptive markers placed over the skin of the VL (as described by Reeves[145]). The acquired images were combined for $ACSA_{VL}$ measurement (I-Figure 12B). The $ACSA_{VL}$ was measured three times using ImageJ and the mean value was recorded for further analysis. The ultrasound scan was made by the same investigator with good test consistency (ICC = 0.99).

Isometric elbow flexion torque (at a 60° elbow flexion) and knee extension torque (at a 60° knee flexion) were measured by a customized dynamometer (MMU, UK) which was calibrated prior to each strength measurement session.



I-Figure 12. Ultrasound images of the upper arm and the thigh. a. Ultrasound image of the upper arm (Sagittal); b. Ultrasound image of the thigh (cross-sectional).

3.2.3. Statistical terminologies in this thesis

Z-score and t-value: a z-score measures the distance between a value and the population mean in the scale of standard deviation. It is calculated as the difference (between a value and the population mean) divided by the standard deviation. A zero z-score means the tested value equals to the population mean. A t-value in a t-test is similar to a z-score, it is used to evaluate if the mean values between two groups are significantly different (i.e. whether the two groups statistically belong to one group). The larger a t-value is, the more different the two group means will be.

Coefficient of determination (R^2), adjusted coefficient of determination (adjusted R^2) and partial coefficient of determination (partial R^2): R^2 is the squared correlation coefficient from a regression model. It represents the proportion of variance in the dependent variable e.g. individual variation in knee extension strength a regression model can explain. It ranges between 0 and 1. Usually, a model with a larger set of independent variables will have a higher R^2 because the dependent variable will be more precisely explained by the model. However, adding more variables might also lead to overfitting, which means the model will fit well with the existing data, but will have a poor performance in predicting future data. To better evaluate the predictive ability of a model, adjusted R^2 is introduced by adjusting the R^2 with the number of predictor variables: adjusted $R^2 = 1-(1- R^2)\times\frac{n-1}{n-(k+1)}$, where n is the sample size and

k is the number of independent variables. Therefore, the adjusted R^2 will increase only when the newly added independent variable can increase the explained variance of the dependent variable to a considerable extent. The partial R^2 represents the proportion of variance of the dependent variable explained by a specific dependent variable in a model.

M and β value: As previously described in the section 1.4.3, a β value is the percentage of methylation in a given CpG site. It has a range of 0-1. An M value is a logit transformation of the ratio of methylated to unmethylated probes. It has an infinite range. The M value is more statistically valid for a differential methylation analysis,[129] and therefore, was used for differential methylation analysis in this thesis (**paper 3**). The beta value is an intuitive biological interpretation. Therefore, it is used for plotting methylation levels in this thesis (**paper 3** and **4**).

4. Supplementary files

This thesis includes many genetic and DNA methylation profiles, and analysis results which are too large to present as appendices. Therefore, some of the data and results are uploaded as supplementary files to a Google drive which is accessible through the following link:
https://drive.google.com/open?id=1T9TVoOGgxYUbl7MzNSeH7kFWJugQqAap

PART 2

RESEARCH PAPERS

CHAPTER 1

Genetic variants and exercise-related muscle changes in older people

Paper 1: Genetic predisposition score predicts the increases of knee strength and muscle mass after one-year exercise in healthy elderly

Paper 2: The genetic effect on muscular changes in an older population: a follow-up study after one-year cessation of structured training

Paper 1

Genetic predisposition score predicts the increases of knee strength and muscle mass after one-year exercise in healthy elderly

Lingxiao He, Evelien Van Roie, An Bogaerts, Christopher I Morse, Christophe Delecluse, Sabine Verschueren, Martine Thomis

Published in *Experimental Gerontology*, 2018; 111: 17-26.

Abstract

This study aims to identify a genetic predisposition score from a set of candidate gene variants that predicts the response to a one-year exercise intervention. 200 participants (aged 60–83 years) were randomly assigned to a fitness (FIT), whole-body vibration (WBV) and control group. Participants in the exercise (FIT and WBV) groups performed a one-year intervention program. Whole-body skeletal muscle mass (SMM) and isometric knee extension strength (PT_{IM60}) were measured before and after the intervention. A set of 170 muscle-related single nucleotide polymorphisms (SNPs) were genotyped. Stepwise regression analysis was applied to select significantly contributing SNPs for baseline and relative change parameters. A data-driven genetic predisposition score (GPS) was calculated by adding up predisposing alleles for each of the phenotypes. GPS was calculated based on 4 to 8 SNPs which were significantly related to the corresponding phenotypes. These SNPs belong to genes that are involved in myoblast differentiation, muscle and bone growth, myofiber contraction, cytokines and DNA methylation. GPS was related to baseline PT_{IM60} and relative changes of SMM and PT_{IM60} in the exercise groups, explaining the variance of the corresponding parameter by 3.2%, 14% and 27%, respectively. Adding one increasing allele in the GPS increased baseline $PTIM_{60}$ by 4.73 Nm, and exercise-induced relative changes of SMM and PT_{IM60} by 1.78% and 3.86% respectively. The identified genetic predisposition scores were positively related to baseline knee extension strength and muscle adaptations to exercise in healthy elderly. These findings provide supportive genetic explanations for high and low responders in exercise-induced muscle adaptations.

Keywords: Exercise; Aging; Genetic predisposition score; Muscle adaptations

1. Introduction

Increasing longevity throughout the world in recent decades has brought healthy aging to the attention of both gerontology and kinesiology researchers. Past studies have found a loss of muscle mass and decrease in muscle performance as two of the most prominent features during the aging process. Such age-associated muscular decline is known as sarcopenia.[37] Using magnetic resonance imaging, Janssen et al. discovered an onset of muscle mass degeneration among subjects in their thirties, with the decay reaching a significant level in the fifth decade [27]. This decrease was mainly caused by the loss of muscle mass in the lower body.[27,146] Similar to muscle mass loss, muscle strength also decreases with aging, but at a

faster rate.[4] This functional weakness is thought to be associated with many factors such as denervation in aged muscle,[33] declined function in mitochondria,[39,147] elevated type I/type II fiber ratio[36,37] and alteration in contractile properties.[41] Consequently, these alterations in muscle morphology and function are closely related to decreased mobility,[21] higher risk of falls,[22] and even increased mortality rate[23] in elderly population.

It is now well established that regular participation in exercise programs can help reduce age-associated functional declines. Multiple exercise protocols have been reported as effective in slowing the muscular aging process. Resistance training and combined aerobic and resistance training have been proven to maintain muscle performance.[43,148,149] For instance, a 26-week exercise intervention in obese elderly found an 18% improvement in strength after combined aerobic and resistance training and a 19% strength increment after resistance training.[44] Muscle power output and muscle maximal strength were also enhanced in response to different training strategies in healthy elderly.[150] Meanwhile, whole-body vibration (WBV) training has been introduced as an alternative for resistance training. Through external vibrations exerted by vibration platform, sensory receptors-mostly muscle spindles-are stimulated. Such process results in the activation of alpha-motor neurons and muscle involuntary contractions.[47] WBV-induced adaptations in aging muscle mass and strength have been reported by several studies.[151,152] Despite the benefits of exercise, muscle strength and mass adaptations after resistance training showed individual response variability (muscle size change ranging from -11 to 30% and leg strength change ranging from -8 to 60%) among subjects, regardless of age and sex.[52] Similar inter-individual ranges in responses are reported for elbow strength in the FAMUSS study.[53] From the findings of previous studies on the relations between inherited characteristics and physical adaptations to exercise, Thomaes et al. found coronary artery disease (CAD) patients with profitable genes had a significantly higher probability to belong to the high responder group (10% highest increases in peakVO_2)[121] and show higher strength gains[123] after a three-month ambulatory supervised exercise training. These findings indicate that genetic factors might be partly responsible for the variance of physical adaptations after training.

Since early reports on exercise capacity-related genes at the end of twentieth century,[153,154] many studies have shown the relation between hereditary characteristics and physical performance.[63] A recent study done by Papadimitriou

et al. on male Caucasian sprinters found that *ACTN3* 577RR carriers had faster sprint time than their homozygous X allele counterparts and 577R allele counted for nearly 1% of sprint time variance.[155] Petr et al. discovered that relative peak power measured by Wingate test (an ergometer-based power test) among elite male ice hockey players was positively related to *PPARA* gene C allele carriers.[156] However, a considerable number of these studies focused merely on one or a limited number of genes. Considering that muscular performance is affected by the combined influences of multiple genes, a new method needs to be applied in order to study the overall effect of multiple gene sequence variants. Candidate-gene based genetic predisposition score (GPS) have gradually been introduced into predicting performance-related phenotypes. Only a limited number of GPSs come from identified sequence variants from Genome Wide Association Studies (GWASs) when exercise-response phenotypes are concerned. The Heritage Family Study provided evidence for GWAS-based predictive GPSs for the responses in submaximal and maximal oxygen uptake after aerobic training.[157] A more general approach is to build a GPS by summing up the number of predisposing alleles that are significantly related to corresponding phenotypes based on candidate-gene association studies. By means of GPS, heritability studies have been able to show the role genetic factors play in the changes of muscular phenotypes following exercise interventions. Through calculating endurance-specific genetic scores, Santiago et al. found higher mean genetic scores in elite endurance athletes compared to controls.[120] Ruiz et al. showed that professional rowers had more preferable genotypes than subjects from the general population.[119] In the studies of exercise interventions on patients with coronary artery disease, Thomaes et al. found data-driven GPS significantly related to increments of peak VO_2[121] and muscular phenotypes[123] after training. To the best of our knowledge, no studies have been performed combining muscle-related genes with GPS to explain baseline muscular phenotypes and exercise-induced muscular changes in a healthy elderly population. Yet, such studies might be helpful in better understanding individual adaptive variance after exercise and can be useful for the design of more individualized exercise regimens in the future.

Therefore, the aim of present study was to assess the predictive power of data-driven GPSs on baseline muscular phenotypes and muscle adaptations to exercise in a healthy elderly population. We hypothesized that elderly people with a higher

GPS have a higher baseline value and greater muscular improvement than those with a lower GPS.

2. Materials and Methods

2.1 Subjects

Elderly people between 60 to 83 years old were recruited from the local communities of the city of Leuven and its surrounding areas. This sample was previously described in the study of Bogaerts et al.,[138] which studied effects of whole-body vibration (WBV) and fitness (FIT) training on muscle strength in elderly. All the subjects went through a series of medical examinations. Exclusion criteria were skeletal, neuromuscular and cardiovascular disorders that may prohibit training and strength-related tests. People with training experience in the past two years were also excluded. This study was approved by the University's Human Ethics Committee in accordance with the Declaration of Helsinki. Informed consent was given by each subject. 200 participants (104 men, 96 women) agreed to provide a blood sample for DNA analyses and their data were analyzed in this study (see flowchart Figure 1).

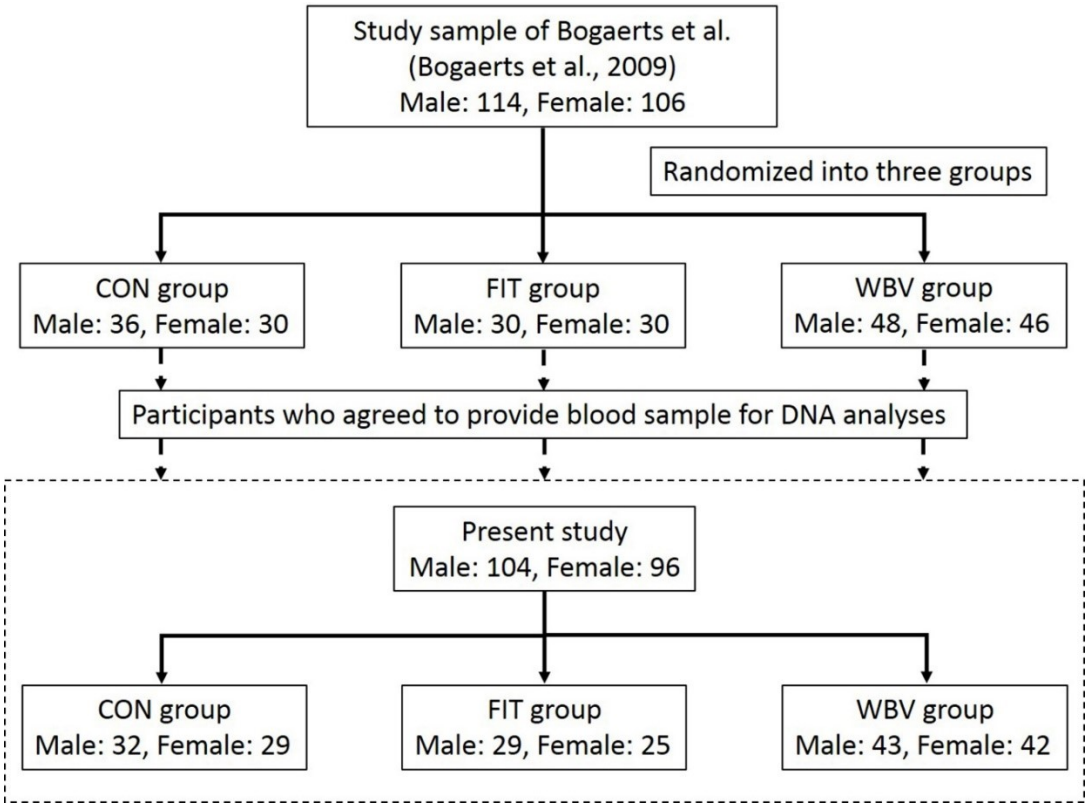


Figure 1. Flowchart of participants in present study.

2.2 Training protocols

Of the 200 participants providing a blood sample, 54 of them performed the fitness program (FIT group), 85 of them were in the WBV group and the rest were control subjects (CON group). Subjects in the FIT and the WBV groups were trained three times a week on nonconsecutive days over a period of one year. All training programs were performed at Leuven University's Training Center under the guidance and supervision of qualified health and fitness instructors.[138]

The training program in the FIT group consisted of aerobic, resistance, balance and flexibility training. It was designed based on the exercise prescriptions for elderly recommended by American College of Sports Medicine (ACSM) guidelines.[158] Subjects firstly performed the aerobic session through one of the four exercises: walking, running, cycling or stepping. The training intensity varied from 70% to 85% of the individual heart rate reserve. The duration of this session was 20 minutes in the starting week and was gradually increased to 45 minutes by the end of the 1-year program. In the resistance training session, subjects performed leg press, leg extension, leg curl (lower body), chest press, upper back, shoulder press, vertical traction, arm curl (upper body), abdominal crunch and back extension (abdominal region) on strength equipment (Technogym Systems, Gambotella, Italy). One repetition maximum (1-RM, the load a participant can only lift once with the best effort) of each exercise for each participant was assessed by qualified health and fitness instructors once a month. The training load started at 50% of 1-RM with 15 repetitions and was gradually increased to 80% of 1-RM with 8 repetitions. 15 minutes of balance exercise and 10 minutes of stretching were performed after each training session. The training programs were described in detail in the study of Bogaerts et al. [138] (P1-Supplementary Table 1).

Participants in the WBV group performed exercises on a vibration platform (Power Plate, Amsterdam, Netherlands) with a maximum duration of 40 minutes. The exercises included body weight squat, deep squat, wide stance squat, toes-stand, toes-stand deep, one-legged squat and lunge. The duration of each exercise started at 30 seconds and was gradually increased to 60 seconds after 9 weeks. A detailed training protocol can also be found in the study of Bogaerts et al.[137] (P1-Supplementary Table 1).

Subjects in the CON group did not undertake any training program. They were advised to maintain their original lifestyle during the study and to not engage in any new physical activity.

2.3 Genotyping

A 4.5 ml blood sample of each participant was collected from an antecubital vein in an EDTA-coated tube. Genomic DNA was extracted using the chemagic Magnetic Separation Module I (chamagic MSM I, PerkinElmer Inc., Waltham, MA, USA) according to the instructions of the manufacturer. Genotyping was done with the Illumina GoldenGate platform (Illumina, Inc., San Diego, CA, USA) at the Genomics Core Facility (UZ/KU Leuven). The selection of genes was based on published articles (up to August 2014) and expression quantitative trait loci (eQTL) analysis. A detailed description of the selection process can be found in the study of Charlier et al.[124] These potential candidate genes were identified for muscular strength or muscular endurance development or regulation. 224 single nucleotide polymorphisms (SNPs) (P1-Supplementary Table 2) came out as muscle-related SNPs. Through blood testing, 12 SNPs were not successfully detected and 3 SNPs presented the same genotypes among all subjects (due to a very low rare allele frequency, P1-Supplementary Table 2). Those 15 SNPs were ruled out from the 224-SNP pool. Results of linkage disequilibrium test showed that 58 SNPs were highly linked as 19 subgroups and one representative SNP was selected from each of these subgroups. A total number of 170 SNPs were withheld for further analyses.

2.4 Muscular phenotype measurements

Whole-body skeletal muscle mass (SMM) was calculated through bioelectrical impedance analysis (BIA). Resistance of BIA was measured by Bodystat 1500MDD (Bodystat Ltd, Douglas, UK) before and after the one-year intervention. Before the test, participants were asked to lie down in a supine position for one minute. During the measurement, two electrodes were placed on the right hand and right foot as instructed in the manual. SMM was calculated for further analyses, using the following regression equation that has been assessed for validity in elderly participants[140]: SMM (kg) = ($Ht^2/R \times 0.401$) + (sex × 3.825) + [age × (-0.071)] + 5.102 where Ht stands for height in centimetres; R stands for BIA resistance in ohms; in sex, men = 1 and women = 0; age is in years.

Biodex Medical System 3 dynamometer (Biodex Company, New York, USA) was used for the measurement of isometric knee extensor strength. This measurement

was done by the same operator before and after the intervention. Before testing, participants were asked to complete a 5-minute warm up on a free-loaded ergometer. Two practice trials were performed to allow for a better understanding of the measuring process. Maximal isometric knee extension strength was evaluated at knee flexion angles of 60° (PT_{IM60} in Nm) with 0° representing full extension.

2.5 Statistical analyses

All the data were reported as mean ± standard deviation (SD) and were analysed using SAS statistical software version 9.4 for Windows (SAS Institute Inc, Cary, NC). Stepwise regression analysis was first used in the detection of SNPs that were significantly related to muscular phenotypes. The significance level to entry was 0.1 and that to stay in the model was 0.05. Alleles that were found positively related to muscular phenotypes from the analysis were regarded as phenotype-related predisposing alleles. Based on the selected significant SNPs from stepwise regression analysis, muscular phenotype-related GPS was calculated with the method used in the calculation of data-driven GPS in the study of Charlier et al.[124] Since the weights of alleles in muscle-related SNPs were not well defined, an accumulative effect was hypothesized and equal weight was given to each predisposing allele. Thus, data-driven GPS of each individual was calculated by adding up all the corresponding predisposing alleles. For example, if the T allele in the rs1130214 of the *AKT1* gene (with T/G alleles) is found to be significantly favourable for ΔPT_{IM60} in the exercise groups, the genotype score of the *AKT1* gene variant is based on the number of T alleles: TT=2, TG=1 and GG=0. The ΔPT_{IM60}-related GPS is then calculated by summing up all genotype scores for SNPs that are found significantly related to PT_{IM60} change. If e.g. 8 SNPs contribute to the GPS, a maximal value of 16 represents the most optimal genetic profile and a value of 0 the worst possible genetic profile (no favourable alleles).

Two-way analysis of variance (ANOVA) was applied to compare between-group values of baseline and one-year relative changes with sex and group as factors. Bonferroni method was used as post-hoc test. Repeated measures ANOVA was used for within-group comparisons of muscular phenotypes between baseline and post-intervention level with sex as a factor. To analyse the effect of GPS on baseline muscular parameters, analysis of covariance (ANCOVA) was performed with age, height, sex and baseline SMM as covariates. To explore the genetic influence on muscle adaptation to exercise, relative changes of muscular phenotypes in exercise

groups were used. The relations between GPS and phenotypic changes after exercise were analysed through ANCOVA with age, height, sex and corresponding baseline muscular value as covariates. P value of 0.05 was set as the level of significance.

3. Results

3.1 Descriptive data

Descriptive data of subjects in each group are presented in table 1. Participants in the three groups had similar age, height and body mass before the intervention. No significant difference in body mass was found among the three groups after one year.

3.2 Baseline muscular phenotypes and training effects

The baseline values and training effects of muscular phenotypes are presented in table 2. At baseline level, SMM and PT_{IM60} showed no significant difference among groups ($p = 0.486$ and $p = 0.805$, respectively). Significant increases of SMM (CON: $p < 0.001$, FIT: $p = 0.006$, WBV: $p = 0.029$) were found in all groups after one year, but these changes among the three groups did not show any differences ($p = 0.299$). After one-year training, PT_{IM60} increased significantly in the two exercise groups (FIT: $p < 0.001$, WBV: $p < 0.001$) while the CON group did not change significantly ($p = 0.744$). Moreover, two-way ANOVA results showed significant differences in relative changes of PT_{IM60} among the three groups ($p < 0.001$). Post-hoc test further showed that the exercise groups increased significantly more than the CON group ($p < 0.05$).

Table 1 Descriptive data of subjects (mean ± SD)

Group	Number	Age (year)	Height (cm)	Body Mass (kg)		
				Pre-intervention	Post-intervention	$\Delta_{post-pre}$ (%)
CON	61	68.23 ± 5.38	167.45 ± 8.54	75.43 ± 10.86	74.49 ± 10.78	-0.98 ± 3.38
FIT	54	67.00 ± 3.88	167.70 ± 9.98	76.13 ± 11.98	74.63 ± 12.19	-1.78 ± 2.99
WBV	85	67.44 ± 4.83	167.22 ± 8.51	75.21 ± 12.62	73.80 ± 11.67	-1.20 ± 3.15
p value (group level)		0.369	1.000	0.958	0.946	0.374

Table 2 Muscular phenotypes before and after one-year intervention (mean ± SD)

Parameter	Baseline	Post-intervention	$\Delta_{post\text{-}baseline}$ (%)
SMM (kg)			
CON	23.68 ± 6.82	24.01 ± 6.09^{+++}	3.96 ± 5.92
FIT	23.65 ± 6.27	24.59 ± 6.65^{++}	3.38 ± 8.06
WBV	23.94 ± 6.50	24.32 ± 6.57^{+}	2.21 ± 6.79
PT_{IM60} (Nm)			
CON	136.29 ± 44.25	138.17 ± 43.51	0.19 ± 16.06
FIT	141.70 ± 39.65	162.43 ± 37.89^{*+++}	14.97 ± 15.57*
WBV	136.92 ± 41.77	151.32 ± 43.47^{+++}	12.09 ± 15.51*

* Significant difference when compared with CON group ($p < 0.05$)

$^{+}$ Significant difference when compared with baseline value ($p < 0.05$)

$^{++}$ Significant difference when compared with baseline value ($p < 0.01$)

$^{+++}$ Significant difference when compared with baseline value ($p < 0.001$)

3.3 Relations between GPS and muscular phenotypes

SNPs closely related to muscular phenotypes were selected through stepwise regression analysis (P1-Supplementary Table 3). Considering that SMM and PT_{IM60} at baseline level were not different among CON, FTI and WBV groups (Table 2), baseline data of the three groups were analyzed together in stepwise regression analysis. Linear relations between GPS and corresponding muscular phenotypes at baseline level are shown in table 3. Since stepwise regression was made separately on each muscular parameter, the number of data-driven SNPs varied with each parameter. As presented in table 3, four SNPs (*ACVR1B*: rs2854464; *FST*: rs3797297; *IGFBP3*: rs3110697; *TTN*: rs10497520) were found significantly related to baseline PT_{IM60}. Data-driven GPS could explain 3.2% of the variance in isometric knee extensor strength. Adding one increasing allele within the GPS increases baseline PT_{IM60} by 4.73 Nm. Results from ANCOVA analysis showed that sex, age and baseline SMM were also significantly related to baseline PT_{IM60}. Although five SNPs (*ACVR1B*: rs2854464; *IGFBP3*: rs3110697, rs6670; *MTRR*: rs327588; *VDR*: rs731236) were found to be closely related to baseline SMM, ANCOVA result did not show a significant relation between baseline SMM and the GPS score (p = 0.250).

Relations between GPS and training responses of SMM and PT_{IM60} in FIT and WBV groups are presented in table 4. Since no significant differences of relative changes were found between FIT and WBV groups (table 2), data in these two exercise groups was analysed together. SNPs closely related to muscle adaptations were selected through stepwise regression analysis (P1-Supplementary Table 3). Six SNPs (*CCL2*: rs4586; *CCR2*: rs768539; *GR/NR3C1*: rs6190; *METTL21C*: rs2390760; *MSTN*: rs2390760; *SPP1*: rs10516796) were found significantly related to SMM changes in the exercise groups. As table 4 shows, GPS, sex, height and baseline SMM were closely related to SMM changes in the exercise groups. Age and training methods (FIT or WBV) did not significantly affect the changes over the one year period. The genotypic predisposition score alone could explain 14% of the adaptive change in SMM and adding one increasing allele to the GPS is associated with a 1.78 % increase in SMM change. Eight SNPs (*AKT1*: rs1130214; *DNMT3L*: rs7354779; *IGFBP3*: rs3110697; *IL15RA*: rs2228059; *MSTN*: rs1805086; *MTRR*: rs162040, rs7703033; *SPP1*: rs10516796) were found to be significantly associated with the change in knee extensor strength by training. The analysis showed that GPS, sex and baseline PT_{IM60} were closely related to PT_{IM60} change in the exercise

groups. GPS alone could explain 27% of the adaptive change. Moreover, increasing the GPS with one predisposing allele is associated with a 3.86% increase in knee extensor strength after training.

Table 3 Relations between genetic predisposition scores and baseline muscular phenotypes

	SMM (kg)				PT_{IM60} (Nm)			
	Estimate	β value	R^2	p	Estimate	β value	R^2	p
GPS	0.17	0.04	0.007	0.250	4.73*	0.12	0.032	0.016
SEX (M=1,F=0)	8.54***	0.66	0.560	<0.0001	18.95*	0.23	0.025	0.034
AGE	-0.06	-0.04	0.011	0.141	-2.01***	-0.23	0.106	<0.0001
HEIGHT	0.23***	0.31	0.235	<0.0001	0.64	0.13	0.017	0.085
$SMM_{baseline}$	-	-	-	-	2.38**	0.37	0.052	0.002
Intercept	-16.01	-	-	-	76.22	-	-	-
Adj. R^2		0.839				0.577		
No. of SNPs		5				4		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Table 4 Relations between genetic predisposition scores and relative changes in muscular phenotypes

	ΔSMM (%)				ΔPT_{IM60} (%)			
	Estimate	β value	R^2	p	Estimate	β value	R^2	p
GPS	1.78***	0.34	0.140	<0.0001	3.86***	0.45	0.270	<0.0001
SEX (M=1,F=0)	10.83***	0.74	0.146	<0.0001	11.53**	0.37	0.110	0.001
EXE (FIT=1,WBV=0)	-0.32	-0.02	0.001	0.770	3.25	0.10	0.022	0.139
AGE	-0.10	-0.06	0.005	0.423	-0.41	-0.12	0.024	0.128
HEIGHT	0.25**	0.32	0.054	0.009	0.22	0.12	0.015	0.232
$SMM_{baseline}$	-1.20***	-1.07	0.217	<0.0001	-	-	-	-
$PT_{IM60\_baseline}$	-	-	-	-	-0.24***	-0.65	0.273	<0.0001
Intercept	-19.54	-	-	-	-3.00	-	-	-
Adj. R^2		0.350				0.511		
No. of SNPs		6				8		

** p < 0.01, *** p < 0.0001

The distribution of GPS and its linear relation with muscular parameters are presented in figure 2 and figure 3. Values for GPS with less than three subjects were pooled together at the lower and upper end of the distribution. As shown in the graphs, at baseline level, subjects with a higher GPS had higher baseline muscle mass and knee extension strength. Higher data-driven GPS is also associated with a larger increment in SMM and PT_{IM60} after one year of exercise training.

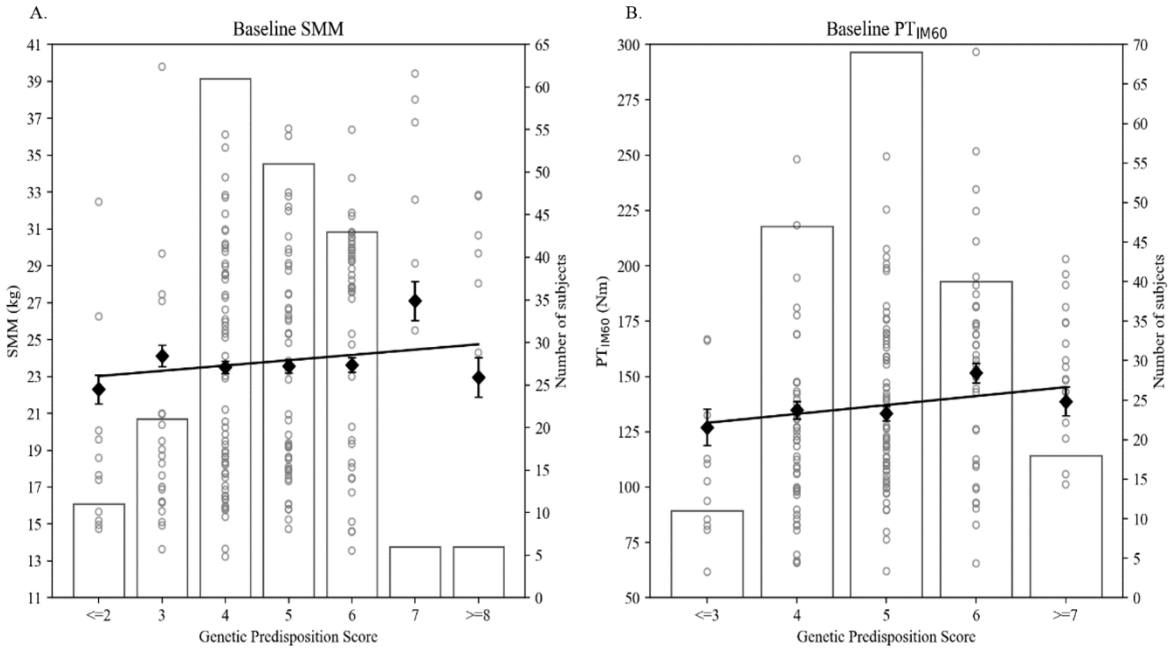


Figure 2. Distribution of GPS and its linear regression model with baseline muscular phenotypes. (A) Linear regression between genetic predisposition score (GPS) and whole-body skeletal muscle mass (SMM) at baseline. GPS is calculated based on 5 SNPs (rs2854464 in ACVR1B, rs3110697 and rs6670 in IGFBP3, rs327588 in MTRR and rs731236 in VDR, n=199). Individual baseline SMM values of the three groups (CON, FIT and WBV) are analysed together and are presented per GPS group on the left y-axis. The trend line shows the relation between GPS and baseline SMM. Least square means of SMM for each GPS is presented as dot with standard errors presented as error bar. Distribution of participants in each GPS is presented in the histogram with number of participants on the right y-axis. (B) Linear regression between GPS and peak isometric knee extension strength at a knee flexion angle of 60° (PTIM60) at baseline. GPS is calculated based on 4 SNPs (rs2854464 in ACVR1B, rs3797297 in FST, rs3110697 in IGFBP3 and rs10497520 in TTN, n=184).

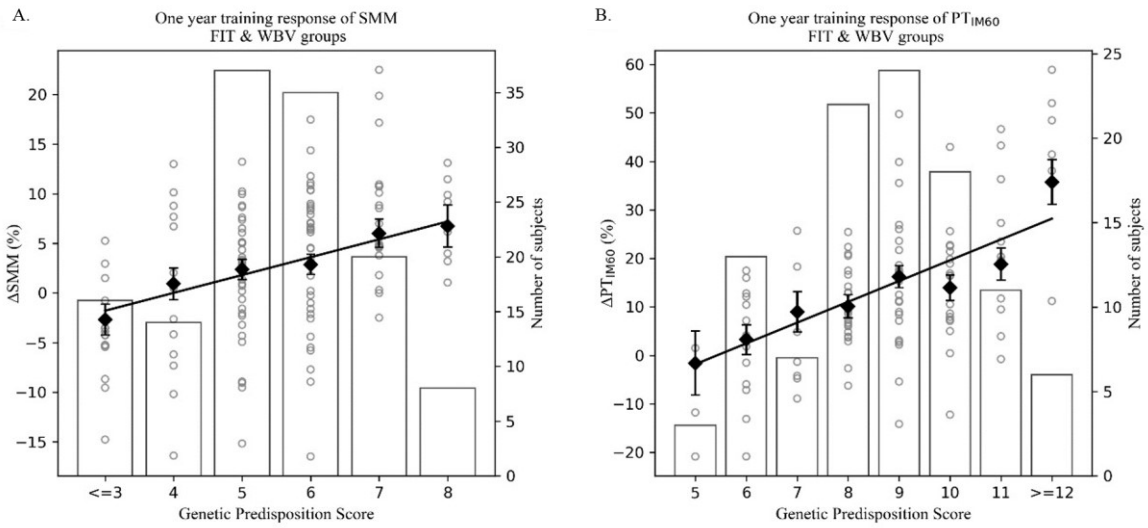


Figure 3. Distribution of GPS and its linear regression model with muscular phenotype changes in exercise groups after one-year training. (A) Linear regression between genetic predisposition score (GPS) and relative changes of skeletal muscle mass (∆SMM) in the exercise groups (FIT and WBV) after one year. GPS is calculated based on 6 SNPs (rs4586 in CCL2, rs768539 in CCR2, rs6190 in NR3C1, rs2390760 in METTL21C, rs3762546 in MSTN and rs10516796 in SPP1, n=130). Individual ∆SMM values is presented on the left y-axis. The trend line shows the relation between GPS and ∆SMM. Least square means of ∆SMM in each GPS is presented as dot with standard errors presented as error bar. Distribution of participants in each GPS is presented in the histogram with number of participants on the right y-axis. Scatterplot is used to present the distribution of SMM change in each GPS group. (B) Linear regression between GPS and relative changes of peak isometric knee extension torque at a knee flexion angle of 60° (∆PT_{IM60}) after one year. GPS is calculated based on 8 SNPs (rs1130214 in AKT1, rs7354779 in DNMT3L, rs3110697 in IGFBP3, rs2228059 in IL15RA, rs1805086 in MSTN, rs162040 and rs7703033 in MTRR, and rs10516796 in SPP1, n=104).

4. Discussion

This study developed GPS to explain the effects of genetic factors on baseline muscular phenotypes and exercise-induced muscular changes in a healthy elderly population. Unlike previous research that studied muscular phenotypes with single or small number of genes, this study was based on 170 SNPs which were selected from a potential set of 224 muscle-related SNPs. Considering the fact that muscular phenotypes are the result of multifactorial and polygenic effects, a larger SNP pool might better explain genetic influences. Although the GPSs were calculated based

on a limited subset of these SNPs, they were positively related to baseline PT_{IM60} and changes of SMM and PT_{IM60} in the exercise groups after one year of training. A genetic predisposition score based on 4 to 8 SNPs explained 0.7 to 3.2 % of variance in baseline and 14 to 27 % in the inter-individual changes in response to training.

Genetic predisposition scores

The specific set-up of this one-year exercise intervention program in elderly subjects and the multi-gene variant approach makes comparisons with other studies difficult. When considering knee extension strength in the untrained state, Thomaes et al.[123] showed no relation between a GPS and isometric knee extension strength in CAD patients, while our study showed that a higher GPS was positively related to a higher **baseline** PT_{IM60}. Although no common SNPs were found between our study and the study of Thomaes et al. in the aspect of knee isometric strength, overlapping SNPs were found in other muscular phenotypes with *GR* rs6190 and *MSTN* rs3762546 closely related to relative change of SMM in our study and relative change of rectus femoris diameter in response to a three-months exercise training in the study of Thomaes et al.[123] A significant relationship between a GPS and isometric knee extension strength was also reported in the study of Charlier et al.[124] With a larger life span sample (200 women, 365 men, 19-73 yr) of adults, a partial r-squared of 5.4% of a GPS based on 8 SNPs was within the same range as 3.2% in our study regarding baseline PT_{IM60}.

Based on the results of ANCOVA in the exercise groups **after training**, GPSs explained 14% of the variance of SMM change and 27% of that in PT_{IM60}. For each increasing allele within the GPS, SMM is predicted to increase by 1.78% and knee extension strength is estimated to gain by 3.86%. Follow-up analysis of the separate training groups (results not shown) indicated that the explained variance for GPSs in the strength response were similar. However, for the response in muscle mass, the GPS in the FIT group had higher predictive value (R^2=30.8%) compared to the WBV group (R^2=6.2%). In the CAREGENE study,[123] the relation between muscular phenotypes and a data-driven GPS based on a 54-SNP pool was studied among CAD patients after a 3-month cardiac rehabilitation training. A GPS based on two SNPs explained 6.25% of the variance in individual responses in isometric knee extension strength. Differences in training programs, duration, SNP pool and subject

characteristics might contribute to the difference in the predictive power of GPSs in both studies.

Which gene variants contribute to the genetic predisposition profiles?

Through stepwise regression analysis (P1-Supplementary Table 3), six genes were found closely related to ***baseline*** SMM and PT_{IM60}. Among these genes, two of them (*ACVR1B*: rs2854464, *IGFBP3*: rs3110697) were associated with both parameters. Windelinckx et al. had found SNP rs2854464 in *ACVR1B* gene to be strongly associated with isometric knee extensor strength with the A-allele as the strength increasing allele.[159] However, no associations between the A-allele and sprint/power performance were also reported in either athletes or control groups from Brazilian[160] and Japanese populations.[161] Based on our data-driven analysis, the G-allele was found predisposed to a higher isometric knee strength and a larger muscle mass. The *IGFBP3* gene was selected into this study because it facilitates myoblast differentiation; specifically the production and secretion of insulin-like growth factor-binding protein 3 (IGFBP3) was in accordance with the differentiation level of myoblast.[162] Rs3110697 in the *IGFBP3* gene was reported as one of the polymorphisms closely related to IGFBP3 blood levels. G-allele carriers were found with higher plasma IGFBP3 level than homozygous A genotypes.[163] This is in line with our finding that the G-allele was positively related to baseline muscular phenotypes. The initial 170 muscle-related SNP pool included several genetic variants in the *FST* gene, which codes for follistatin. Acting as an inhibitor of the myostatin receptor,[164] the overexpression of follistatin could cause dramatic increases in muscle growth.[165] Previously, sex-specific fat free mass was found to be associated with sequence variation in the *FST* gene.[166] We found a relation between rs3797297 and baseline PT_{IM60} with the T-allele as the predisposing allele. *TTN* gene polymorphisms were associated with skeletal muscle fascicle length and marathon performance in habitually trained men.[167] Unlike the study of Thomaes et al.,[123] which failed to find any relation between rs10497520 from *TTN* gene and isometric knee extension strength in CAD patients, our results identified the C-allele of rs10497520 as a predisposing allele of isometric knee extension strength. Finally, the *VDR* gene codes for vitamin D receptor, which plays an important role in calcium homeostasis and muscle function.[168] The Rs731236 variant in the *VDR* gene was associated with hand grip strength.[169] Inconsistent with the finding of Windelinckx et al.,[79] which showed a sex-specific relation between *VDR* polymorphisms and knee extension strength, our result found no significance between *VDR* gene and

isometric knee extension strength, however the gene variant contributed to the GPS for skeletal muscle mass. The sixth variant was rs327588 within the *MTRR* gene that encodes for methionine synthase reductase. This enzyme with DNA methylation-related function is discussed in more detail below.

Training responses on SMM and knee extension strength were found related to 11 genes. Specifically, the *MSTN* and *MTRR* gene contributed two SNPs while other genes only contributed one. The *MSTN* gene encodes myostatin, a protein which negatively regulates the growth of muscle cells. Myostatin deficient mice were found to have larger muscle mass, more type IIB fibres and lower relative force generation ability than wild types.[170] In humans, the R allele of the rs1805086 variant has been associated with lower explosive strength,[77] but increased odds of being a centenarian.[66] AKT, also known as protein kinase B (PKB), is a critical regulator of muscle growth through the IGF1-AKT/PKB pathway.[171] Insulin-like growth factor 1 (IGF-1) was found able to induce myotube hypertrophy through the activation of the AKT pathway.[172] Activation of the *AKT1* transgene in mice also revealed a hypertrophy of type IIB fibres and a counteraction of lean muscle mass loss in aged mice.[173] These results support our finding that *AKT1* gene was related to knee strength gains. The presence of *CCL2* and *CCR2* gene in the GPS for adaptive changes rather than baseline values supported the idea that these two genes were more related to muscle adaptations. CCL2 is expressed by macrophages and muscle satellite cells, its expression is dramatically increased following muscle damage. CCR2 is the receptor of CCL2. Previous studies have found that the expressions of both genes were associated with muscle exercise-induced damage and the speed of recovery, which varied with individuals.[174,175] *NR3C1* polymorphisms have been reported related to many sex-specific body composition and muscular phenotypes.[176] Recently, it has been shown that *NR3C1* polymorphisms (rs10482614, rs10482616 and rs4634384) were associated with muscle strength and size response after a 3-month resistance training.[177] Our results showed another SNP (rs6190) in the *NR3C1* gene to be associated with knee extension strength changes after training. Methyltransferase like 21C (METTL21C) not only participated in protein-lysine methyltransferase activity but was found to affect bone and muscle metabolism as well.[178] Hangelbroek et al. found that higher expression of the *METTL21C* gene was associated with frailty status in both young and elderly subjects[179] while we found this gene was related to exercise-induced SMM change. The A-allele frequency in *IL15RA* rs2228059 was

higher than C-allele in cyclists while the opposite direction of A/C allele frequencies was found in elite rowers and triathletes.[180] The A-allele in rs2228059 was also reported associated with larger muscle volume but lower muscle quality in men.[181] In our study, rs2228059 was only related to knee strength adaptation after training. A study on Duchenne muscular dystrophy patients showed the *SPP1* gene as a determinant of this disease with G-allele carriers in SNP rs28357094 suffering from a more rapid degenerating progress.[182] Although that SNP was also included in our initial SNP pool, rs10516796 came out as the only SNP in the *SPP1* gene that showed close relation with muscular changes after exercise.

Noticeably, through stepwise regression, three variants (rs162040, rs327588 and rs7703033) in the *MTRR* gene were identified related to baseline SMM or one-year PT_{IM60} response. The *MTRR* gene expresses methionine synthase reductase which participates in the metabolic cycle that provides methyl groups to DNA.[183] A/G heterozygotes and G homozygotes of the rs1801394 variant in *MTRR* gene were found more frequently in athletes when compared with non-athletes, indicating a reduced DNA methylation capacity might be induced by systematic training.[184] Considering the reports of the *MTRR* gene affecting muscular metabolism through DNA methylation,[184,185] we hypothesize that DNA methylation may contribute to the variability of muscle adaptations induced by exercise. Subjects with more predisposing alleles of the *MTRR* gene (C allele in rs162040 and G allele in rs7703033) may trigger a larger extent of DNA hypomethylation in the *MTRR* gene region which leads to an upregulation in myogenic proteins[184] after one year training, resulting in a higher improvement of knee extension strength. We also observed an association between the *DNMT3L* gene variant rs7354779 and knee strength change after exercise. Suetake et al. observed in mice that DNA (cytosine-5)-methyltransferase 3-like (DNMT3L) plays a crucial role in the activation of DNA (cytosine-5)-methyltransferase 3A (DNMT3a) and 3B (DNMT3b), two major DNA methyltransferases responsible for the creation of DNA methylation patterns.[186]

Limitations and future directions

It should be acknowledged that compared to more stringent measures of SMM (e.g. using DEXA, MRI), the standard error of estimate value was 2.7kg (9%) in the application of BIA-estimated muscle mass.[140] Greater associations may have been observed had a more accurate measure of SMM been adopted. The mild increases in SMM (2-3% increment in exercise groups) might also be related to the combined

training design. Although a high resistance load of 70-80% 1RM was used in this study, the 2.2-3.4% one-year gain of SMM was similar to the 2.6-3.2% increase of muscle volume in the study of Van Roie et al. who only designed a 12-week pure resistance intervention for the elderly with the same load.[187] Moreover, in the calculation of GPS, each predisposing allele was given equal weight. This ignored the fact that every genetic variant might contribute differently to muscular phenotypes. Other GPS calculation methods, such as total weighting genotype score,[188] LASSO and Elastic Nets[189] can provide new ways to study the relation between sets of gene variants and aging muscle. Noticeably, through stepwise regression in the selection of data-driven SNPs, this study failed to identify some genes which were previously reported to be associated with muscle and power performance, such as rs1815739 in *ACTN3* gene[155] and rs4253778 in *PPARA* gene.[156] This might be due to the interaction with other SNPs or be related to the fact that the *ACTN3* R allele is more strongly related to dynamic contractions at high velocities[155] compared to maximal isometric strength as was measured in this study. GWAS-identified gene variants would certainly strengthen the set of candidate-gene based variants to build genetic predisposition profiles for trainability phenotypes.[190] SNPs identified in a recent GWAS study for grip strength would have provided additional loci beyond the SNP pool in this study.[191] Given that our sample was limited in size, the GPS-construction phase and test for predictive value was done in the same set of subjects. The predictive value of the GPS should therefore be tested in an independent study with similar subject and exercise intervention characteristics in a first phase, and in other exercise interventions or subject characteristics in future studies. The one-year intervention design of the study limited the sample size to 104 (200 for baseline values). A power of 80% (at alpha=0.05) can be reached with effect sizes of 0.26-0.30 (ANOVA test for a single SNP) or explained variances of 4-8% (GPS for regression). A larger sample would therefore be able to include more SNPs with smaller effect sizes in the GPS, which could explain the inter-individual differences in muscular fitness and responses to training to a larger extend.

The focus of this study was on the role of a genetic profile based on candidate gene variants for muscle mass and strength phenotypes and responses to training. However, gene variants related to aerobic phenotypes (e.g. peakVO_2) might be different depending on selected gene sets and on the specific mode of training.[192] This indicates that genes associated with muscle adaptations in our study might not

be a robust predictor of adaptations in aerobic capacity from the same population. Subjects with limited strength responses might be high-responders in aerobic parameters in response to this mixed-type intervention. We explored this by adapting the GPSs (based on muscle strength/mass related phenotypes) to the change in peakVO_2 that was also available in these subjects and hypothesized a negative relationship. However, the results showed that the GPSs for ΔSMM and ΔPTIM60 were not significantly related to the aerobic adaptation in this sample (p = 0.23, p = 0.49, respectively) with estimated coefficients of 1.13 and -0.48. It would be more optimal to genotype GWAs-identified SNPs and candidate-gene variants related to aerobic performance and adaptations and to explore specific GPSs for these aerobic fitness response phenotypes.

Figures 2 and 3 of the GPS distribution and its linear regression with muscular parameters also showed individual variability among subjects within the same GPS group. The smallest standard error of muscular changes was 1.02% in ΔSMM under the GPS group of 6 while the largest standard error was found as 6.61% in ΔPT_{IM60} with the GPS group of 5. Such findings imply that other unknown exercise-related genes are involved in the process and it suggests that genetic composition is not the only factor to affect muscular training responses. In fact, the expression of a gene can also be affected without the alteration of genetic sequence, this process is known as epigenetics.[193] Many external factors, such as nutrient intake, activity level and living environment can contribute to the modification of DNA (de-)methylation.[194] The involvement of sequence variants in *MTRR*, *DNMT3L* and *METTL21C* genes discussed above also suggested the existence of epigenetics in training adaptation processes. Furthermore, Barrès et al. have reported a global hypomethylation and reduced methylation levels in promoter regions of energy metabolism related genes (*PGC-1α*, *PDK4* and *PPAR-δ*) in human muscle biopsy after an acute peak pulmonary oxygen uptake rate test.[195] Thus, further research on the relation between epigenetic factors and aging muscle and its responses to exercise is needed.

This study only focused on genetic effects, which is a relatively popular field in the study of muscular training responses. To strengthen the understanding of genomics in effects of exercise in elderly, further research on the epigenome, and integration of transcriptomics, proteomics and metabolomics are needed.[196,197] Muscular phenotypes adjusted for daily physical activity and nutrient intake could make the conclusions more convincing as well.

Conclusion

In conclusion, based on a 170 muscle-related SNP pool, we found that a data-driven GPS was positively related to baseline isometric knee strength and adaptive changes of muscle mass and knee extensor strength after one-year exercise in a healthy elderly population. Specifically, the GPS explained part of the inter-individual variance of training response with some DNA methylation-related genes involved in the adaptive process. These findings provide additional genetic explanations for individual differences in exercise-induced changes. Further research into the role of DNA methylation effects on training adaptations are also suggested.

5. Conflict of interest

The authors declare no conflicts of interest.

6. Funding

This study was supported by the Flemish Government within the First Generation Policy Research Centre for Sports, Physical Activity and Health, the Scientific Research Fund - Flanders (FWO) [grant G-0521-05] and a PhD grant to Lingxiao He by the Erasmus Mundus ‘Move-AGE’ joint doctorate program.

Paper 2

The genetic effect on muscular changes in an older population: a follow-up study after one-year cessation of structured training

Lingxiao He, Evelien Van Roie, An Bogaerts, Christopher I Morse, Sabine Verschueren, Christophe Delecluse, Martine Thomis

Prepared for *Experimental Gerontology*.

Abstract

Purpose: To explore the genetic effect on changes of muscle mass and muscle strength one year after the cessation of a structured training intervention in an older population using data-driven genetic predisposition scores (GPSs).

Methods: Participants (n = 113, aged between 61-81 yr) who performed one-year of combined fitness (n = 44) or whole-body vibration (n = 69) training were reassessed one year after the cessation of the training. Whole-body skeletal muscle mass, isometric knee extension strength, isotonic knee extension velocity, and isokinetic knee extension strength at 60°/s and at 240°/s were assessed. From a set of 170 muscle-related single nucleotide polymorphisms (SNPs), stepwise regression analysis was applied to select favourable SNPs that were significantly related to changes in muscular phenotypes. To analyse the overall genetic effect, data-driven GPSs were calculated by summing up predisposing alleles, and were evaluated in a general linear model with sex, age, body mass index and post-training values of skeletal muscle mass or muscle strength as covariates.

Results: 46 SNPs from 34 genes were identified as being significantly linked to one-year cessation-related muscular alterations. These genes cover the domains of DNA methylation, metabolism, muscle growth, muscle structure and neural control. Data-driven GPSs and one-year cessation-related muscular alterations were significantly related ($p < 0.01$). Participants with higher GPSs had less muscular declines during the cessation period while data-driven GPSs accounted for 26–37% of the phenotypic variances.

Conclusions: Our findings provide supportive evidence for a genetic association with muscular changes after the cessation of a structured training among older people, indicating that the maintenance of training benefits is partially gene-related.

Keywords: genetic predisposition score, muscle, older adults, cessation of structured training

1. Introduction

The process of ageing is commonly accompanied by progressive loss in skeletal muscle mass and muscle strength.[198] A 3-year follow-up study of Goodpaster et al.[23] on older adults aged 70–79 years has revealed a 1% annual loss in leg lean mass with muscle strength decreasing three times faster than muscle mass. These declines in knee extensor strength and thigh muscle mass are associated with

increased risk of mobility loss in the older population.[21] Consequently, physical performance and quality of life in older adults are largely affected by functional and structural alterations in ageing muscles.[81]

Exercise has been well established as an effective non-pharmacological method to counteract muscle degeneration in older adults. Resistance training has long been suggested as an effective intervention for frail older adults.[199] It helps to increase maximal muscle strength and muscle mass, and to reduce risk of falls to a considerable extent.[199,200] Other exercise interventions, such as whole-body vibration (WBV) training and combined training (consisting of both aerobic and resistance exercises), are also reported as effective in ameliorating ageing muscle conditions. A previous review of WBV training effect in elderly has shown that WBV may improve isometric knee strength, muscle power and balance control to a similar extent as traditional resistance training.[201] Meanwhile, combined training (four times per week) carried out among men aged 40–67 years was also reported to exert a similar improvement in maximum leg extension strength (with an average increase of 22%) as that induced by a resistance training (twice per week, with an average increase of 21%).[202]

Besides the large number of studies on exercise benefits for the ageing muscle, many researchers have also focused on the lasting benefits of exercise by describing the loss of muscle strength and size following exercise cessation, termed as “detraining”. For example, the lasting of training benefits is training intensity-dependent. In the study of Fatouros et al.,[203] strength and mobility gains of older men who received a high intensity resistance training lasted longer than those who trained at a low intensity. Moreover, muscle size and muscle strength do not decrease at the same speed during detraining. Older women who completed 12 weeks of resistance training retained a 12% gain in knee extensor strength after 3-months detraining, while the muscle volume of knee extensors had already dropped back to baseline levels.[204] These muscular decreases during detraining are multifactorial. Composition changes such as fat infiltration[205] and reduced cross-sectional area of type I and type II fibers[206], morphological alterations like decreased pennation angle and fascicle length[207], neural control[208] and hormone[209] changes have all been reported to be associated with decreased muscle strength in the detraining period. However, the role of genetics in determining the rate of muscular changes after the cessation of structured training is poorly understood. It is known that some of the variances within the training response can be attributed to genetic

factors in young adults[210] and older women[211]. As reported in the study of Delmonico et al.,[211] older women with *ACTN3* R577 XX genotype had higher baseline knee extensor power than R-homozygous carriers, while the latter had greater improvement after 10 weeks of strength training. A greater understanding of genetic impact on the variances of muscle mass and muscle strength during the cessation period could be through considering multiple favourable genotypes in constructing genetic predisposing scores (GPS). The GPS has been adopted to explain variances in thigh muscle mass and knee strength after cardiac rehabilitation training,[123] to explore muscular changes with ageing,[124] and to evaluate athletic status.[119] Given the recent findings that GPS explained 14% and 27% of exercised induced increases in muscle mass and muscle strength,[212] we hypothesise a genetic effect on muscular changes after the cessation of a structured training intervention. Therefore, the purpose of this research is to study the genetic effect on muscular changes after one-year of exercise cessation in an older population.

2. Methods

2.1 Participants

Participants, aged between 61 and 81 years, were originally recruited in an exercise intervention study of Bogaerts et al.[138] Older adults with physical disorders that might affect exercise performance or with any training experience in the past two years were excluded. In the study, participants were randomly assigned into a control (CON) group, a combined fitness (FIT) group or a WBV group (Figure 1). The training intervention lasted for one year and the participants were not aware of a follow-up test. One year after the training program, participants who had provided blood samples for genotyping in the exercise (FIT or WBV) groups, were contacted for a follow-up test. Since our previous study has reported the adaptive changes of muscle mass and muscle strength induced by exercise,[212] this study mainly focused on the muscular changes in these exercise groups after the cessation. Noticeably, the study of Bogaerts et al.[138] only included the participants who fully completed the training program. In our study, we included participants who had more than 60% of attendance during the training and completed at least one of the follow-up measurements (Figure 1). This ensured a comparatively large sample size for further genetic study. This study was approved by the University Ethics Committee and all the participants were asked to sign an informed consent form.

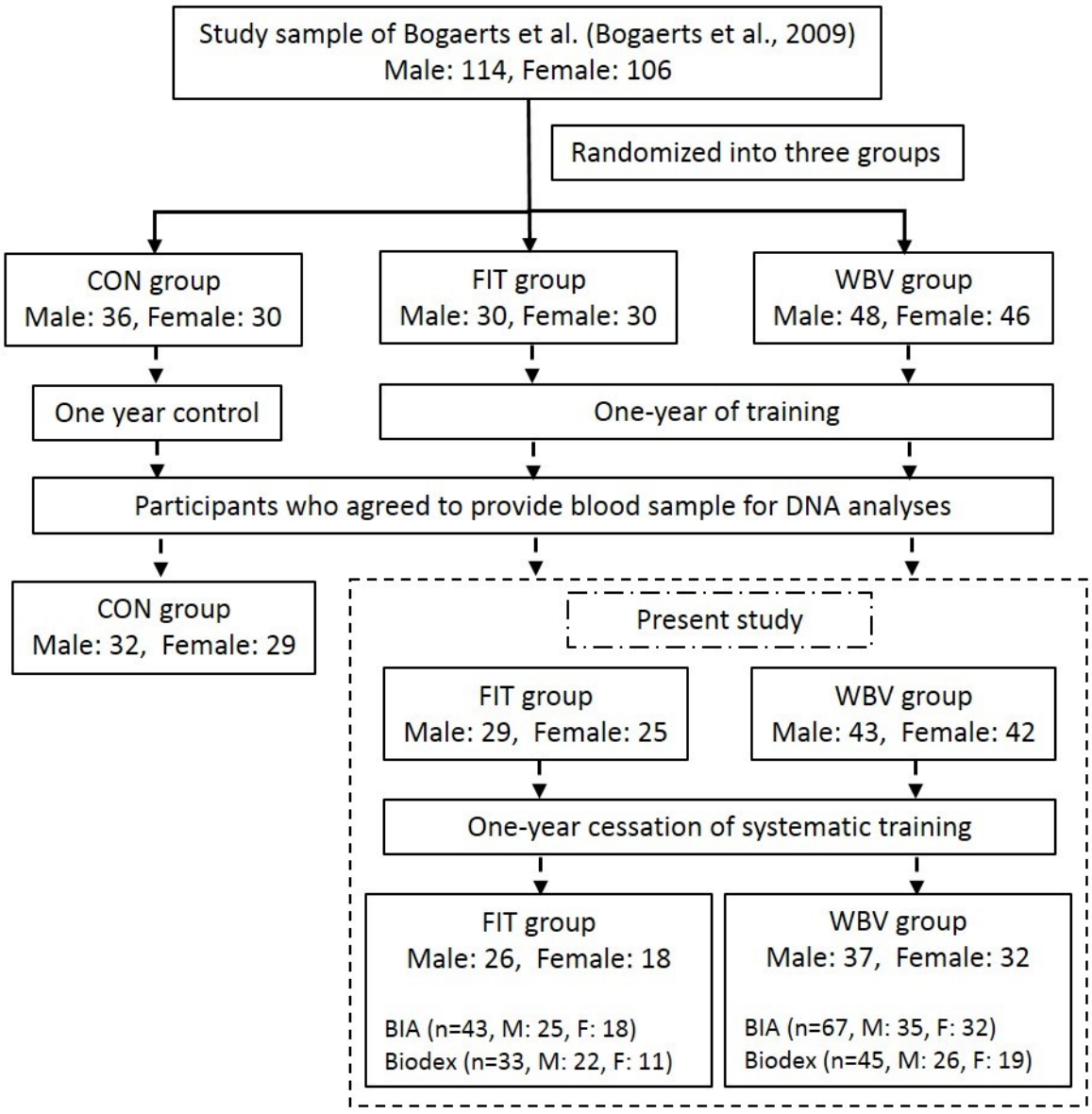


Figure 1. Flowchart of participants in present study.

2.2 Training protocols

The training programs (P2-Supplementary Table 1) have been described in detail in the study of Bogaerts et al.[138] Briefly, participants in the exercise groups trained three times per week on non-consecutive days for one year. The training program for the FIT group was designed following the ACSM guidelines for older adults exercise prescription,[158] which consisted of aerobic, resistance, balance and flexibility exercises. Participants in the WBV group were instructed to perform static and dynamic leg exercises on vibration platforms (Power Plate, Amsterdam, Netherlands). The training programs were performed at Leuven University's Training Center under the guidance and supervision of qualified health and fitness instructors. Participants in the CON group were advised to maintain their lifestyle and to not engage in any new physical activity.

2.3 Genotyping

A 4.5 ml venous blood sample was collected from each participant using an EDTA-coated tube. DNA was extracted using the chemagic Magnetic Separation Module I (chemagic MSM I, PerkinElmer Inc., Waltham, MA, USA). Genotyping was completed with GoldenGate assay (Illumina, Inc., San Diego, CA, USA) following the protocols of the manufacturer.[213] Single nucleotide polymorphisms (SNPs) that were reported to be associated with the development or regulation of muscle function or muscle growth were selected based on published articles (up to August 2014) and expression quantitative trait loci (eQTL) analysis. In total, 224 muscle-related SNPs (P2-Supplementary Table 2) were genotyped from each blood sample. From the genotyping results, 12 SNPs had a detection success rate of less than 80%; 3 SNPs showed the same genotypes among all the participants; 58 SNPs had high linkage disequilibrium (absolute correlation coefficient greater than 0.8) within 19 subgroups. Within these subgroups, the SNP with the largest number of correlated SNPs or published references was selected as a representative for each subgroup. In the end, 54 SNPs were excluded from the initial SNP pool and 170 SNPs were kept for further analyses.

2.4 Parameter measurements

Electrical resistance of the body was measured by bioelectrical impedance analysis (BIA) using Bodystat 1500MDD (Bodystat Ltd, Douglas, UK). Skeletal muscle mass (SMM) was estimated using the following equation which was developed by Janssen et al.[140]: SM mass (kg) = ($Ht^2/R \times 0.401$) + (sex × 3.825) + [age × (-0.071)] + 5.102 where Ht stands for height in centimetres; R stands for BIA resistance in ohms; in sex, men = 1 and women = 0; age is in years. SMM calculated by this equation showed validity among older adults with a standard error of estimate of 2.7kg (9%).[140]

Isometric, isotonic and isokinetic knee extensor strength was tested using Biodex Medical System 3 dynamometer (Biodex Company, New York, USA). Participants were asked to complete a 5-minute warm up on a free-loaded cycle ergometer followed by two practice trials on the dynamometer to ensure some familiarization. In the actual tests, each of the following protocols were performed twice and the maximum value of each protocol was recorded for further analyses.

Isometric test: Peak torque of isometric knee extension was measured at a knee flexion angle of 60° (PT_{IM60} in Nm, 0° representing full extension) with a duration of

5 seconds. Maximal isometric strength at the flexion angle of 90° was also recorded for load setting in the isotonic test.

Isotonic test: The isotonic test included 3 sets of ballistic knee extension movements with a load of 20% of the peak isometric strength obtained at the knee flexion angle of 90°. Starting at the knee flexion angle of 90°, participants were asked to extend their legs as fast as possible until they achieved the knee flexion angle of 20°. Peak velocity (PV_{IT20} in °/s) was recorded for further analyses.

Isokinetic test: Participants performed isokinetic knee extension and flexion movements at two different speeds. The first measurement required participants to complete four repetitions at a low velocity of 60°/s. The second measurement consisted of six repetitions at a higher velocity of 240°/s. Peak torque of knee extensors at 60°/s (PT_{IK60} in Nm) and at 240°/s (PT_{IK240} in Nm) were recorded and further analysed.

2.5 Statistical analyses

All data are reported as mean ± standard deviation (SD) and were analysed using SAS statistical software version 9.4 for Windows (SAS Institute Inc, Cary, NC). Since muscle mass and muscle strength can be affected by multiple factors, the effect of a single gene on muscle is rather limited. Therefore, an accumulative effect of multiple gene variants was hypothesized in this study. Similar to the data-driven method used in the study of Charlier et al.,[124] alleles that were positively related to muscular changes were regarded as predisposing alleles and were equally weighted as 1. Stepwise regression analysis, with an entry/exit significance of 0.1/0.05, was used in the selection of SNPs (from a SNP pool of 170) that were significantly related to relative change of each muscular phenotype after the cessation. Genetic predisposition score (GPS) of each participant was calculated by adding up the weight of each phenotype-driven genotype. For example, using stepwise regression, allele G of SNP rs3762546 in gene *MSTN* was found to be favourable for ΔPT_{IM60} after one-year of cessation. Thus, the genotype score of rs3762546 was calculated based on the number of G allele: GG=2, CG=1 and CC=0. ΔPT_{IM60}-driven GPS in a participant was calculated by summing up scores of all the SNPs that were found significantly related to corresponding phenotypes.

Comparisons between the FIT and the WBV groups at post-training and one-year follow-up tests were made by two-way analysis of variance (ANOVA) with sex and group as factors. Bonferroni method was applied as post-hoc test. The same

ANOVA was also completed in the comparisons of relative changes of muscular phenotypes after one-year of exercise cessation. To compare the value of each muscular phenotype between post-training and follow-up tests, repeated measures ANOVA was made with sex and group as factors. A *p* value of 0.05 was set as the level of significance. The association between GPS and relative changes of muscular parameters was evaluated by general linear model (GLM) with age, sex, body mass index (BMI) and corresponding post-training muscle values as covariates.

3. Results

3.1 Descriptive data and relative changes at post-training and follow-up tests

Descriptive data of muscular phenotypes in the FIT and the WBV groups are presented in Table 1. Between-group comparisons showed that participants in the FIT and the WBV groups were not different for muscle mass and muscle strength at both post-training and follow-up tests ($p > 0.05$). By comparisons between post-training and one-year follow-up test, significant increases in BMI were found for both exercise groups ($p < 0.01$) one year after the cessation of structured training. Moreover, PV_{IT20} ($p < 0.01$), PT_{IK60} ($p = 0.02$) and PT_{IK240} ($p < 0.01$) decreased significantly in both exercise groups. Time*sex, time*group or time*sex*group interactions were non-significant for all phenotypes. Table 2 presents the number of participants with increased/decreased muscle mass and strength after the one-year cessation. The majority of participants had decreased dynamic strength (i.e. PV_{IT20} PT_{IK60} and PT_{IK240}) while approximately half of the participants experienced decreases in SMM and PT_{IM60}.

Table 1 Descriptive data and p values from ANOVA of between group comparisons at post-training and follow-up tests

Parameters		Post-training	Follow-up	$\Delta_{Follow-Post}$ (%)	*p* values from repeated measures ANOVA			
					Time	Time*Sex	Time*Group	Time*Sex*Group
AGE (year)					-	-	-	-
FIT								
	F	66.44 ± 3.79	-	-				
	M	67.48 ± 3.96	-	-				
WBV								
	F	67.07 ± 5.17	-	-				
	M	67.79 ± 4.51	-	-				
p value at Group level		0.55						
p value at Group*Sex level		0.84						
Height (m)					-	-	-	-
FIT								
	F	160.02 ± 7.90	-	-				
	M	174.32 ± 6.09	-	-				

WBV							
F	161.22 ± 5.67	-	-				
M	173.08 ± 6.50	-	-				
p value at Group level	0.99						
p value at Group*Sex level	0.28						
Body mass (kg)				<0.01**	0.19	0.84	0.49
FIT							
F	66.62 ± 9.39	66.32 ± 8.92	-1.36 ± 2.80				
M	82.04 ± 9.57	83.15 ± 9.51	-0.09 ± 2.90				
WBV							
F	68.65 ± 9.27	68.58 ± 8.86	-0.17 ± 3.53				
M	78.95 ± 11.64	80.21 ± 12.77	-0.44 ± 3.21				
p value at Group level	0.77	0.85	0.97				
p value at Group*Sex level	0.15	0.16	0.61				

power at Group level	0.06	0.05	0.05				
power at Group*Sex level	0.30	0.29	0.08				
BMI (kg/m^2)				<0.01**	0.28	0.90	0.47
FIT							
F	26.08 ± 3.86	26.20 ± 3.83	-1.36 ± 2.80				
M	27.13 ± 3.32	27.42 ± 3.35	-0.09 ± 2.90				
WBV							
F	26.44 ± 3.50	26.47 ± 3.44	-0.17 ± 3.53				
M	26.43 ± 3.62	26.62 ± 3.63	-0.44 ± 3.21				
p value at Group level	0.79	0.68	0.97				
p value at Group*Sex level	0.41	0.39	0.61				
power at Group level	0.06	0.07	0.05				
power at Group*Sex level	0.13	0.14	0.08				

SMM (kg)				0.45	0.96	0.55	0.83
FIT							
F	18.04 ± 2.00	17.71 ± 2.34	1.40 ± 8.29				
M	30.21 ± 3.04	29.99 ± 3.05	4.21 ± 6.28				
WBV							
F	18.46 ± 2.18	18.49 ± 2.61	2.76 ± 9.52				
M	30.33 ± 3.23	30.84 ± 5.58	4.32 ± 17.25				
p value at Group level	0.58	0.29	0.53				
p value at Group*Sex level	0.76	0.97	0.95				
power at Group level	0.09	0.19	0.10				
power at Group*Sex level	0.06	0.05	0.05				
PT_{IM60} (Nm)				0.43	0.93	0.64	0.64
FIT							
F	127.92 ± 18.18	127.79 ± 26.66	13.43 ± 17.70				

M	186.32 ± 28.17	186.63 ± 32.58	16.50 ± 17.73				
WBV							
F	123.05 ± 27.56	125.63 ± 24.80	15.32 ± 18.18				
M	181.48 ± 36.61	174.20 ± 37.29	6.79 ± 22.37				
p value at Group level	0.41	0.31	0.76				
p value at Group*Sex level	1.00	0.48	0.64				
power at Group level	0.13	0.17	0.06				
power at Group*Sex level	0.05	0.11	0.08				
PV_{IT20} (°/s)				<0.01**	0.39	0.68	0.67
FIT							
F	330.17 ± 37.73	307.58 ± 58.96	-1.63 ± 11.56				
M	377.62 ± 34.91	353.95 ± 35.29	-1.68 ± 9.37				
WBV							

F	328.08 ± 31.45	321.75 ± 33.10	0.87 ± 12.18				
M	364.79 ± 36.99	345.05 ± 40.26	-0.34 ± 15.02				
p value at Group level	0. 29	0.78	0.65				
p value at Group*Sex level	0.45	0.22	0.85				
power at Group level	0.18	0.06	0.07				
power at Group*Sex level	0.12	0.24	0.05				
PT_{IK60} (Nm)				0.02*	0.25	0.27	0.56
FIT							
F	111.78 ± 17.98	102.65 ± 25.28	2.09 ± 6.79				
M	168.54 ± 29.57	164.18 ± 30.00	5.70 ± 13.44				
WBV							
F	106.50 ± 18.50	107.98 ± 18.23	0.72 ± 8.9				
M	158.26 ± 28.67	156.29 ± 33.40	0.29 ± 17.16				
p value at Group level	0.12	0.84	0.14				

p value at Group*Sex level	0.61	0.30	0.71				
power at Group level	0.35	0.06	0.31				
power at Group*Sex level	0.08	0.18	0.07				
PT_{IK240} (Nm)				<0.01**	0.97	0.50	0.85
FIT							
F	60.46 ± 10.26	53.11 ± 15.94	-0.83 ± 8.22				
M	93.58 ± 16.04	89.14 ± 14.63	3.76 ± 14.76				
WBV							
F	57.54 ± 10.45	57.24 ± 10.06	3.17 ± 10.19				
M	85.64 ± 14.62	82.28 ± 14.52	0.54 ± 16.63				
p value at Group level	0.04	0.66	0.32				
p value at Group*Sex level	0.34	0.08	0.51				
power at Group level	0.53	0.07	0.17				

power at Group*Sex level	0.16	0.41	0.10

* $p < 0.05$, ** $p < 0.01$

Table 2 Count of participants with increased/decreased muscle mass and strength after one-year cessation of structured training

Muscular phenotype	No. of participants with percentage change <= 0	No. of participants with percentage change > 0	Percentage of participants with negative percentage change (%)
SMM (kg)			
FIT	19	24	44
WBV	34	33	51
PT_{IM60} (Nm)			
FIT	15	18	45
WBV	23	22	51
PV_{IT20} (°/s)			
FIT	27	5	84
WBV	33	11	75

PT_{IK60} (Nm)			
FIT	25	8	76
WBV	24	21	53
PT_{IK240} (Nm)			
FIT	22	11	67
WBV	23	11	68

Table 3. Regressions of data-driven GPSs and relative muscular changes after one-year cessation of structured training

	GPS	SEX (M=1,F=0)	AGE	BMI	Corresponding post-training value	Intercept	Adj. R^2	No. of SNPs
ΔSMM (%)								
Estimate	2.09	-0.91	0.07	0.18	-			
β value	0.52	-0.07	0.05	0.09	-			
Partial R^2	0.27	0.01	<0.01	0.01	-	-29.36	0.27	9
p	<0.01	0.39	0.58	0.27	-			
Δ PT_{IM60} (%)								
Estimate	4.53	3.02	-0.02	0.50	-0.06	-38.69	0.32	7

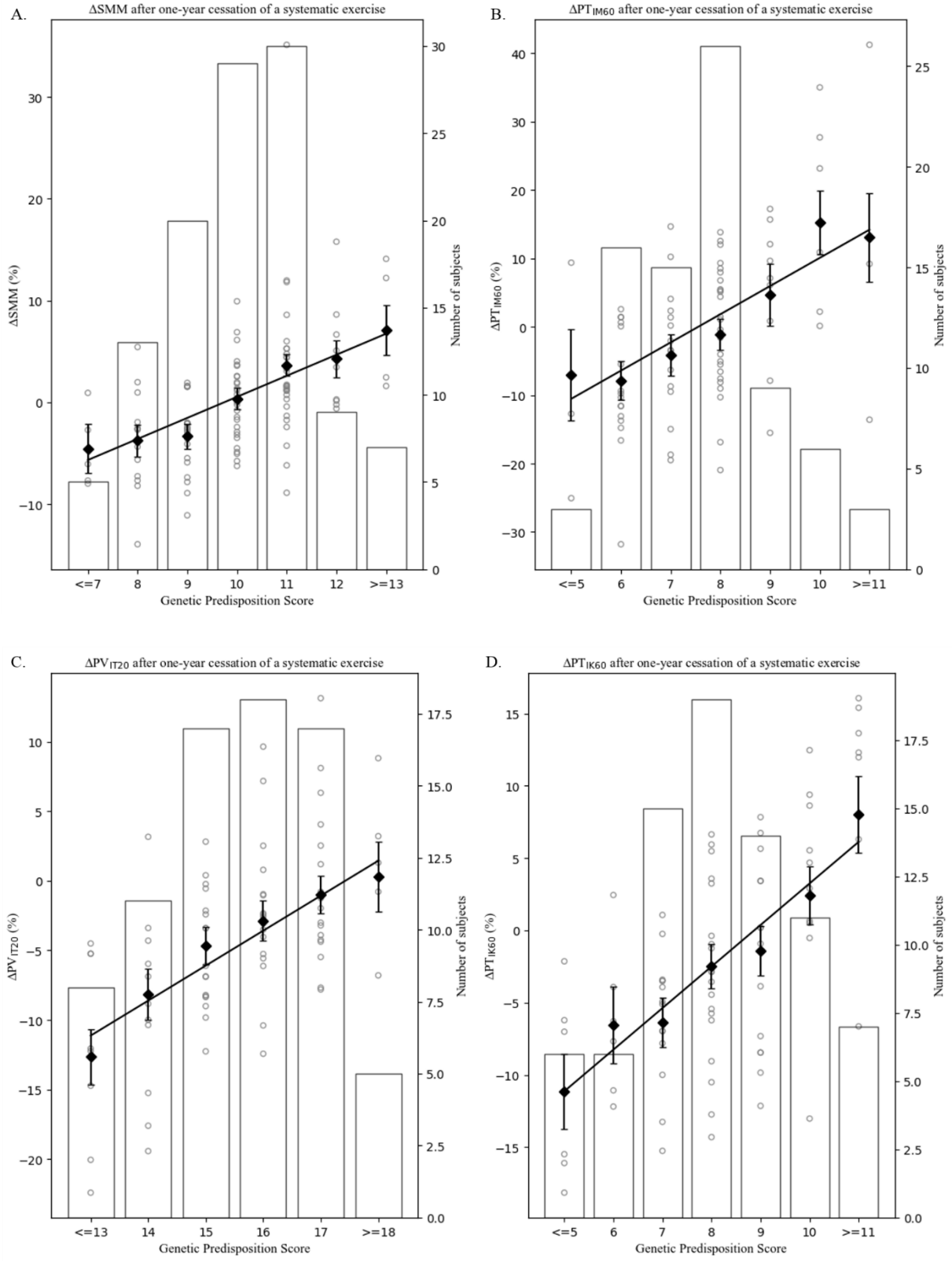
β value	0.53	0.11	-0.01	0.12	-0.20			
Partial R^2	0.27	0.01	<0.01	0.02	0.03			
p	<0.01	0.45	0.96	0.22	0.18			
Δ PV_{IT20} (%)								
Estimate	2.24	1.90	-0.31	-0.09	-0.04	-3.40	0.40	13
β value	0.59	0.14	-0.18	-0.04	-0.22			
Partial R^2	0.36	0.02	0.04	<0.01	0.06			
p	<0.01	0.22	0.08	0.66	0.05			
Δ PT_{IK60} (%)								
Estimate	2.74	2.39	-0.16	0.23	-0.01	-19.44	0.37	9
β value	0.62	0.15	-0.08	0.09	-0.04			
Partial R^2	0.37	0.01	0.01	0.01	<0.01			
p	<0.01	0.31	0.42	0.33	0.76			
Δ PT_{IK240} (%)								
Estimate	2.56	0.84	-0.03	0.34	0.02	-68.75	0.27	18
β value	0.52	0.05	-0.01	0.12	0.04			
Partial R^2	0.26	<0.01	<0.01	0.02	<0.01			
p	<0.01	0.78	0.90	0.23	0.78			

3.2 Associations of GPS with relative muscular changes after one-year cessation of structured training

Since no significant differences were found in relative changes between the FIT and the WBV groups, values of the two groups were analysed together for the selection of data-driven SNPs and the evaluation of genetic influence on muscular changes after the one-year cessation of structured training. Muscular phenotype-driven SNPs are presented in detail in P2-Supplementary Table 3, in which we showed that unlike many genes that contributed only one SNP to muscular changes, more than one SNP was identified in gene *ACVR1B*, *ATP1A2*, *MTHFR* and *MTRR*, respectively. Furthermore, rs2251375 in *H19*, rs3741211 in *IGF2*, rs2390760 in *METTL21C*, rs3762546 in *MSTN*, rs1805087 in *MTR*, rs327575 and rs97713 in *MTRR*, and rs4790881 in *SMG6* were found to be linked with more than one change in muscular parameters (P2-Supplementary Table 3). Yet, no SNP was found to be associated will all the muscular parameters.

GPS was calculated by summing up the weight of predisposing SNPs. The results of GLM are presented in Table 3. These results showed that data-driven GPS was closely associated with changes in muscular phenotypes one year after the cessation of a structured training ($p < 0.01$). Noticeably, GPS accounted for similar variances (from 26% to 37%) in muscle mass and muscle strength changes during the cessation period. Increasing the data-driven GPS with one predisposing allele is associated with an increase from 2.09% to 4.53% in the change of SMM, PT_{IM60}, PV_{IT20}, PT_{IK60} and PT_{IK240}. Since muscle strength decreased after the exercise cessation, results from table 3 indicate that participants with higher GPS had less loss in muscle strength than lower GPS carriers.

GPS distribution of participants and linear models between GPS and over time changes of muscular phenotypes are presented in Figure 2. GPS was categorized with no less than 3 participants in each group. As shown in Figures 2a-e, participants with higher GPS had less decreases in muscle mass and muscle strength after one-year cessation of a structured training.


A.
ΔSMM after one-year cessation of a systematic exercise
ΔSMM (%)
Number of subjects
Genetic Predisposition Score
<=7
8
9
10
11
12
>=13
B.
ΔPTIM60 after one-year cessation of a systematic exercise
ΔPTIM60 (%)
Number of subjects
Genetic Predisposition Score
<=5
6
7
8
9
10
>=11
C.
ΔPVIT20 after one-year cessation of a systematic exercise
ΔPVIT20 (%)
Number of subjects
Genetic Predisposition Score
<=13
14
15
16
17
>=18
D.
ΔPTIK60 after one-year cessation of a systematic exercise
ΔPTIK60 (%)
Number of subjects
Genetic Predisposition Score
<=5
6
7
8
9
10
>=11

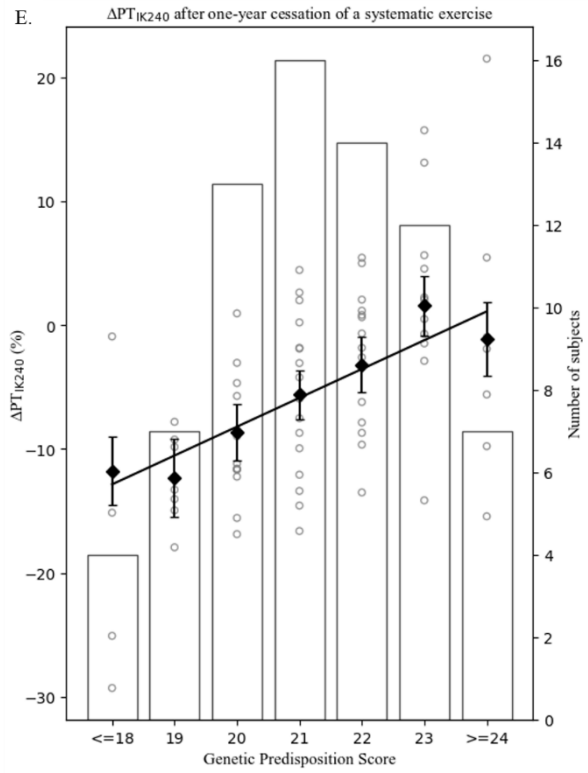


Figure 2. Distribution of GPS and its linear regression model with muscular phenotype changes after one-year cessation of a structured training. (A) Linear regression between genetic predisposition score (GPS) and relative change of skeletal muscle mass (∆SMM) in the exercise groups (FIT and WBV) after one year cessation of a structured training (adjusted for age, sex, BMI and corresponding post-training value). GPS is calculated based on 9 SNPs from 9 genes (rs4870044 in ESR1, rs11549465 in HIF1A, rs3741211 in IGF2, rs7924316 in IGF2AS, rs2390760 in METTL21C, rs3762546 in MSTN and rs97713 in MTRR, rs2229139 in RYR1, and rs4790881 in SMG6). Individual ∆SMM values is presented on the left y-axis. The trend line shows the relation between GPS and ∆SMM. Least square means of ∆SMM in each GPS is presented as dot with standard errors presented as error bar. Distribution of participants in each GPS is presented in the histogram with number of participants on the right y-axis. Scatterplot is used to present the distribution of ∆SMM in each GPS group. (B) Linear regression between GPS and relative change of peak isometric knee extension torque at a knee flexion angle of 60° (∆PTIM60) after one year cessation of a structured training (adjusted for age, sex, BMI and corresponding post-training value). GPS is calculated based on 7 SNPs from 7 genes (rs2296383 in CACNA1S, rs8111989 in CKM, rs689 in INS, rs2390760 in METTL21C, rs3762546 in MSTN, rs327575 in MTRR, and rs28357094 in SPP1). (C) Linear regression between GPS and relative change of peak velocity of isotonic knee extension (PV_{IT20}) after one year cessation of a

structured training (adjusted for age, sex, BMI and corresponding post-training value). GPS is calculated based on 13 SNPs from 11 genes (rs3733890 in BHMT, rs6107853 in BMP2, rs1800169 in CNTF, rs4511463 in GSC, rs2251375 in H19, rs3741211 in IGF2, rs11121828 in MTHFR, rs1805087 in MTR, rs97713, rs1801394 and rs162031 in MTRR, rs1800470 in TGFB1, and rs1483246 in ZNF804A). (D) Linear regression between GPS and relative change of peak torque of isokinetic knee extension at 60°/s (PT_{IK60}) after one year cessation of a structured training (adjusted for age, sex, BMI and corresponding post-training value). GPS is calculated based on 9 SNPs from 8 genes (rs2854248 in ATP1A2, rs10883631 in FN1, rs17727841 in IGF1, rs2390760 in METTL21C, rs1801133 in MTHFR, rs327575 and rs7703033 in MTRR, rs4790881 in SMG6, and rs10497520 in TTN). (E) Linear regression between GPS and relative change of peak torque of isokinetic knee extension at 240°/s (PT_{IK240}) after one year cessation of a structured training (adjusted for age, sex, BMI and corresponding post-training value). GPS is calculated based on 18 SNPs from 14 genes (rs746434 and rs10783485 in ACVR1B, rs12721026 in APOA1, rs1016732 in ATP1A2, rs3797297 in FST, rs2251375 in H19, rs2919358 in KBTBD13, rs1137101 in LEPR , rs3762546 in MSTN, rs1476413 and rs1009592 in MTHFR, rs1805087 in MTR, rs10475399, rs326123 and rs9313211 in MTRR, rs4950877 in MYOG, rs4253778 in PPARa, and rs142196418 in RIMS1).

4. Discussion

4.1 Are gene variants related to muscular changes after the cessation of a structured training?

Using the methods of stepwise regression and data-driven GPS, this study analysed the overall genetic effect on muscular changes after one-year cessation of a structured training in an older group. From a 170-SNP pool, 46 SNPs of 32 genes (P2-Supplementary Table 3) were found to be closely associated with muscular changes. GLM results showed that participants with higher GPSs (more favourable alleles) are less likely to lose muscle mass and muscle strength after the cessation of training. Based on these models, data-driven GPSs explained 26–37% of the variances of these muscular changes during the cessation.

The set-up of multi-gene variants and an exercise cessation background makes it difficult to compare our results with other studies. To our knowledge, there is presently no research among older adults regarding the genetic influence on muscular changes following a cessation of training, with limited research

investigating the genetic influence on muscular adaptations resulting from exercise intervention. A cross-sectional study carried out by Charlier et al.[124] among 565 Flemish Caucasians (aged 19–73 yr) showed that 4.6–6.6% of variances in muscle mass and muscle strength could be explained by data-driven GPS. Such limited degrees of explainable variance by GPS might be due to the wide age range in which many non-genetic factors can affect muscular phenotypes in the long term. Therefore, when restricting the set-up to a shorter age range, an increased role for GPS (as what we have found in this study) can be observed. The degree of genetic variation contributing to muscular changes after the cessation of exercise (26–37%) are similar to those reported for responses to exercise interventions. With a set of 54 SNPs, data-driven GPS-explained 6–26% of variances in knee extension strength and muscle size adaptations after a 3-month training among coronary artery patients.[123] Our previous study in the same study population also found that data-driven GPS accounted for 14% and 27% of the variances in ΔSMM and ΔPT_{IM60}, respectively, after a one-year exercise intervention.[212]

In addition, the present study found a few SNPs that were previously reported to be associated with exercise-induced muscular gains. Some of those SNPs even contributed to the change in the same phenotype. Based on our results, rs1016732 from gene *ATP1A2* showed an association with the decreased PT_{IK240} after training cessation while it also contributed to the increased peak torque of knee extension at a high speed of 180°/s in response to a cardiac rehabilitation program.[123] Similarly, another SNP (rs2854248) from gene *ATP1A2* showed association with ΔPT_{IK60} both in our study and that of Thomaes et al.[123] Since the favourable alleles from SNPs in the study of Thomaes et al.[123] were not presented, we could not make comparisons regarding to the direction of each SNP. Furthermore, He et al.[212] reported that *METTL21C* rs2390760 (with C as the favourable allele) and *MSTN* rs3762546 (with G as the favourable allele) were significantly related to increased muscle mass (ΔSMM) after WBV and FIT training while these SNPs were also closely associated with the one-year cessation-related SMM change in our study. However, in the present study, allele G was found as a favourable allele in SNP rs2390760 and allele G remained as the favourable allele in SNP rs3762546. This suggests that carriers of the C allele in *METTL21C* rs2390760 are more susceptible to exercise than G allele carriers while allele G in *MSTN* rs3762546 is predisposing for the adaption of muscle mass in exercise as well as its maintenance after the cessation

4.2 What kind of genes are related to muscular alterations after the cessation of exercise?

Although the validation on datasets with other older adults still remains to be tested, our findings suggest some representative variants out of a large SNP set that are significantly related to muscular changes after exercise cessation. Based on the categories in P2-Supplementary Table 2, among the genes that had significant associations with muscular changes in this study, 3 genes are involved in DNA methylation, 3 genes are related to hormone expression or its receptor, 9 genes encode for growth/differentiation factors, 9 genes are metabolism-related, 7 genes contribute to muscle/bone structure and 3 genes are involved in neural control.

The discovery of a contribution of gene *MTHFR*, *MTR* and *MTRR* to muscular changes indicates the involvement of DNA methylation after the cessation of exercise. DNA methylation is one of the mechanisms in epigenetic processes, which regulates gene expression without entailing a change in the DNA sequence.[83] Generally, hypermethylation in promoter regions will repress transcriptions of corresponding genes while hypomethylation will reactivate them. Recent studies have shown that methylation changes can be induced by exercise. In the study of Barrès et al.,[195] muscle biopsies were collected 20 mins after an acute aerobic capacity test and hypomethylation were found in promoter regions of several metabolism-related genes (*PGC-1α*, *PDK4* and *PPAR-δ*). Meanwhile, hypomethylation also took place in some genes (*BICC1*, *STAG1*, *GRIK2* and *TRAF1*) after both a single bout and a 7-wk resistance training program, and returned to baseline levels after a cessation of 7-wk.[102] In our present study, we found that genetic variation in *MTHFR*, *MTR* and *MTRR* genes, which encode for corresponding enzymes that regulate the methylation circle,[214] might play a role in altered methylation during the cessation period. Therefore, it is likely that a DNA hypomethylation favourable gene might be associated with a better response towards training as well as a longer maintenance of the gains when a structured training stops.

Genes related to hormone expression, muscle growth/differentiation, metabolism or muscle/bone structure have been linked to physical performances by many studies.[65,66,210] *PPARa* intron 7 (rs4253778) G/C polymorphism has been reported as exercise-oriented with a high frequency (80%) of GG genotype existing among endurance athletes. Further biopsy analysis showed a higher percentage of slow-

twitch fibers in GG carriers when compared with the CC counterpart.[215] Similarly, in the aspect of one-year cessation-related muscular changes among the participants in our study, allele G was found favourable (with less decrease) for the change of dynamic muscle strength at a high contraction speed (ΔPV_{IK240}).

In the domain of neural factors, this study identified three SNPs from three genes, among which is the gene *CNTF*. Encoding for ciliary neurotrophic factor, the rs1800169 polymorphism in gene *CNTF* has been found to be associated with muscle strength in several studies. Walsh et al. reported a sex-specific effect in gene *CNTF* G/A polymorphism with only women of homozygous G alleles improving more in isometric elbow strength than A-allele carries after a 12-wk upper arm training.[216] This is consistent with our results of the association between rs1800169 and dynamic knee contraction performance under a low load (ΔPV_{IT20}) with allele G exerting a favourable effect after the cessation of exercise. Yet, our findings contradict the cross-sectional study of De Mars et al.,[217] who studied 493 adults (aged 38-80 yr) and found that polymorphisms in gene *CNTFR* rather than *CNTF* were related to knee extension strength differences.

4.3 Are the genes associated with one-year cessation-related muscular changes the same as those related to a long term muscular ageing process?

Generally, the one-year cessation of exercise in our study can be regarded as a one-year ageing process. Therefore, we compared our results with other genetic studies on a long term muscular changes related to ageing. A 5-year longitudinal study carried out by Delmonico et al.[218] among older adults aged 70-79 years found no significant association between *ACTN3* R577X (rs1815739) polymorphism and declined muscle strength in ageing. Another longitudinal study (mean follow-up period: 14.2 yr) made by Schrager et al.[219] also showed that the *IGF2* ApaI (rs680) polymorphism was not related to losses of arm endurance capacity and grip strength. Similarly, these genes did not show significant linkage to muscular changes in our study. Yet, we did find 7 common gene variants (P2-Supplementary Table 4), which were favourable for one-year exercise cessation-related muscular changes in the present study, that were previously reported in a cross-sectional study on muscle mass and muscle strength among adults with an age range of 19-73 years[124] despite that the favourable alleles in some of these genes are not consistent with that in the present study.

4.4 Limitations

The data-driven GPS is only one approach to investigate the association between one-year exercise cessation-related muscular decreases and gene variants. Many other processing methods such as total GPS, weighted GPS or elastic net GPS were also used in different studies with varying predictive powers. As found in the study of Charlier et al.,[124] elastic net GPS had the best prediction on SMM while data-driven GPS and total GPS had the best prediction on strength-related phenotypes. Based on 6 genetic polymorphisms, Massidda et al.[220] found the weighted GPS explained more in variance of explosive performances (18% squat jump and 24% counter-movement jump) than the total GPS. Noticeably, as presented in the first part of this discussion, data-driven GPS exerted similar predictive power regarding the muscular changes among the older participants after exercise training.[123,212] Therefore, for the consistency of approach in our previous study, we used a data-driven GPS approach in the present study.

Our conclusions are also limited considering the fact that the selection of data-driven SNPs and the predictive power of GPS were tested on the same sample. An application to an independent sample or cross-validation should better testify our findings. Furthermore, although we find that participants with higher muscle-related GPS scores tend to have smaller losses in muscle mass and strength after the cessation of exercise, the result is weakened by the limited number of participants in the highest and lowest GPS groups. As illustrated in Figure 2, a large variance in muscular phenotypes can be found in GPS groups at both ends of the GPS distribution. Therefore, with a larger sample size, there might be more participants at both ends, resulting in smaller standard error of estimates and confidence intervals in these two GPS groups.

Despite that the majority of participants demonstrated decreased dynamic strength (Table 2), this study failed to control for external factors such as food consumption and exercise habits during the one-year of cessation. Perhaps these factors can partially explain the increase of SMM and PT_{IM60} in half of our participants besides the possible explanation of individual variability.

5. Conclusion

In this study, we applied stepwise regression and data-driven GPS methods from a 170-SNP set to explore the genetic effect on decreases of muscular phenotypes after one-year cessation of a structured training. We found that GPSs accounted for

26–37% of the variances of corresponding muscular changes while participants with more favourable gene variants tended to have less declines in those changes. Moreover, 46 SNPs from 34 genes were identified to be significantly associated with these muscular alterations. These genes contribute to the domains of DNA methylation, metabolism, muscle growth, muscle structure and neural control. In addition, our results provide supportive explanations for the involvement of genetic variants in inter-individual variations of the loss of muscular benefits after the cessation of a structured training among the older population.

6. Conflict of interest

The authors declare no conflicts of interest.

7. Acknowledgements

The authors would like to thank the Flemish Government within the First Generation Policy Research Centre for Sports, Physical Activity and Health, the Research Foundation - Flanders (FWO) [grant G-0521-05] and the Erasmus Mundus ‘Move-AGE’ joint doctorate program for their funding supports to this study.

CHAPTER 2

Sarcopenia-related DNA methylation differences

Paper 3: Differentially methylated gene patterns between age-matched sarcopenic and non-sarcopenic women

Paper 3

Differentially methylated gene patterns between age-matched sarcopenic and non-sarcopenic women

Lingxiao He, Praval Khanal, Christopher I Morse, Alun Williams, Martine Thomis

Published in *Journal of Cachexia, Sarcopenia and Muscle*. 2019 Sep 11. DOI: 10.1002/jcsm.12478

Abstract

Background:

Sarcopenia is characterized by progressive decreases in muscle mass, muscle strength and muscle function with ageing. Although many studies have investigated the mechanisms of sarcopenia, its connection with epigenetic factors, such as DNA methylation, still remains poorly understood. The aim of this study was to explore sarcopenia-related DNA methylation differences in blood samples between age-matched sarcopenic and non-sarcopenic older women.

Methods:

A sarcopenic group (n = 24) was identified and selected from a set of 247 older Caucasian women (aged 65–80 yr) based on cut-off points of skeletal muscle index at 6.75 kg/m^2 and grip strength at 26 kg (the lower quintile of grip strength in the set). A non-sarcopenic group (n = 24) was created with a similar age distribution as that of the sarcopenic group. DNA methylation patterns of whole blood samples from both groups were analysed using Infinium MethylationEPIC BeadChip arrays. Differentially methylated CpG sites (dmCpGs) were identified at a *p* value threshold of 0.01 by comparing methylation levels between the sarcopenic and non-sarcopenic groups at each CpG site. dmCpG-related genes were annotated based on homo sapiens hg19 genome build. The functions of these genes were further examined by gene ontology and KEGG pathway enrichment analysis.

Results:

The global methylation level of all analysed CpG sites (n = 788,074) showed no significant difference between the sarcopenic and non-sarcopenic groups (p = 0.812), while the average methylation level of dmCpGs (n = 6,258) was significantly lower in the sarcopenic group (p = 0.004). The sarcopenic group had significantly higher methylation levels in TSS200 (the region from transcription start site to 200 nucleotides upstream of the site) and lower methylation levels in gene body and 3'UTR regions. In respect of CpG regions, CpG islands in promoters and some intragenic regions showed greater levels of methylation in the sarcopenic group. dmCpG-related KEGG pathways were mainly associated with muscle function, actin cytoskeleton regulation and energy metabolism. Seven genes (*HSPB1*, *PBX4*, *CNKSR3*, *ORMDL3*, *MIR10A*, *ZNF619* and *CRADD*) were found with the same methylation direction as previous studies of blood sample methylation during ageing.

54 out of 4,335 genes were shared with previous studies of resistance training.

Conclusion:

Our results improve understanding of epigenetic mechanisms of sarcopenia by identifying sarcopenia-related DNA methylation differences in blood samples of older women. These methylation differences suggest underlying alterations of gene expression and pathway function, which can partially explain sarcopenia-related muscular changes.

Keywords: Sarcopenia, Older women, DNA methylation, Differentially methylated CpG sites, Pathway analysis

1. Introduction

DNA methylation is a mechanism of regulation of gene expression without alterating the original gene sequences.[83] In mammals, cytosine is the most common base where methylation takes place.[221] Methylation of cytosine involves the attachment of a methyl group to the 5' position of cytosine and can be found in 57–85% of cytosin-phosphate-guanine (CpG) sites.[88] Most CpG sites scatter in mammal genomes; yet, there are regions with clustered CpG sites, known as CpG islands,[222] which can be found in 72% of gene promoters.[89] The dynamic change of DNA methylation is connected to the regulation of gene expression during development and differentiation.[91] Methylated CpG islands in gene promoters have been associated with long-term gene silencing.[223] Moreover, methylated CpG islands of intragenic regions have been found to influence various functions; for example, intragenic DNA methylation in transcriptionally active genes can impede gene expression by reducing elongation efficiency of RNA polymerase II.[93] Methylation in intragenic regions might also prevent gene bodies from spurious transcriptions,[224] and the activities of some methylated intragenic CpG islands are possibly regulated by other CpG islands acting as initiators of transcription.[94] Besides CpG islands, the methylation of CpG shores (sequences within 2 kb distance from CpG islands[225]) is also found in the regulation of gene expression.[226] DNA methylation patterns can be modified by many factors such as age, air polution, lifestyle, nutrition and training.[102,104,105,195,227,228] The association between ageing and DNA methylation has been studied in various tissues such as saliva,[100] blood,[229,230] muscle,[101] skin[231] and brain.[232] In vitro myoblast cultivation demonstrated that an acute early proliferative lifespan TNF-α exposure induced a long-term maintenance of elevated myoD methylation, indicating an underlying epigenetic regulation that might be

related to muscle loss in later life.[233] Zykovich et al. identified 500 ageing-related CpGs as possible predictors of chronological/biological age by comparing DNA methylation patterns in skeletal muscle biopsies between old and young adults.[101] Besides ageing studies, DNA methylation changes have also been related to many disease conditions such as breast cancer,[226] rhabdomyosarcoma[234] and juvenile dermatomyositis.[235]

Sarcopenia has been recognised as a muscle disease, which is characterized by progressive decreases in muscle mass and muscle function.[6] Although ageing is the primary factor, other factors, such as disuse and malnutrition, have also been identified as covariates of sarcopenia.[9] Considering that these factors are also reported to be associated with methylation changes, a possible relationship between sarcopenia and DNA methylation is suggested. Notably, sarcopenia-related changes are not restricted to the muscle itself, as endocrine disorders are also linked to sarcopenia.[135] Inflammatory cytokines such as tumour necrosis factor (TNF)-α and interleukin-6 (IL-6) increase muscle loss and impair muscle regenerating capacity during the ageing process.[134] Serum levels of IL-6, secreted protein acidic and rich in cysteine (SPARC) and macrophage migration inhibitory factor (MIF) were found to be higher in those with sarcopenia compared to controls while insulin-like growth factor 1 (IGF-1) level was significantly lower in sarcopenics.[136] A combined score on these serum levels could be used as a biomarker for sarcopenia,[136] therefore a specific focus on DNA methylation in blood could add to the knowledge of more systemic factors contributing to sarcopenia. If DNA methylation in blood (partially) overlap with those from muscle biopsy based DNA, the more easily accessible venous blood samples can be studied in further research and (biomarker) applications.

The present study aimed to analyse whole blood-based methylation differences between sarcopenic and non-sarcopenic populations. The results of this study will enrich our understanding of sarcopenia by identifying differentially methylated CpG (dmCpG) sites and possible alterations in related gene expression and corresponding signalling pathways.

2. Methods

2.1. Participants

A set of 247 older, independently living, Caucasian women (aged 65–80 yr) were recruited from the local area, and provided written informed consent following local

ethics approval (Manchester Metropolitan University, Crewe, UK). Of the initial 247, 168 provided 5 mL venous blood samples. These were subsequently categorised into sarcopenic (n = 25) and non-sarcopenic (n = 138) groups using cut-off points of skeletal muscle index (SMI, calculated using skeletal muscle mass divided by height squared) at 6.75 kg/m^2 [16] and hand grip strength (HGS) at 26 kg (the lower quintile of HGS in the recruited set). Through a process of further selection including age matching,[98] completeness of data, rankings of SMI and hand grip strength z score, and summed z score (Figure 1), 24 participants (age of sarcopenic group 72.5 ± 4.2 yr, non-sarcopenic group 70.5 ± 3.3 yr) from each group were selected for DNA methylation analysis. In the sarcopenic group, 21 participants with negative z scores in SMI and HGS were selected first, with an additional three selected via an ascending sequence of summed z scores. Selection in the non-sarcopenic group was done in an opposite direction: 23 participants with positive z scores in SMI and HGS were selected first, with an additional participant with the highest summed z score selected from the remainder (Figure 2).

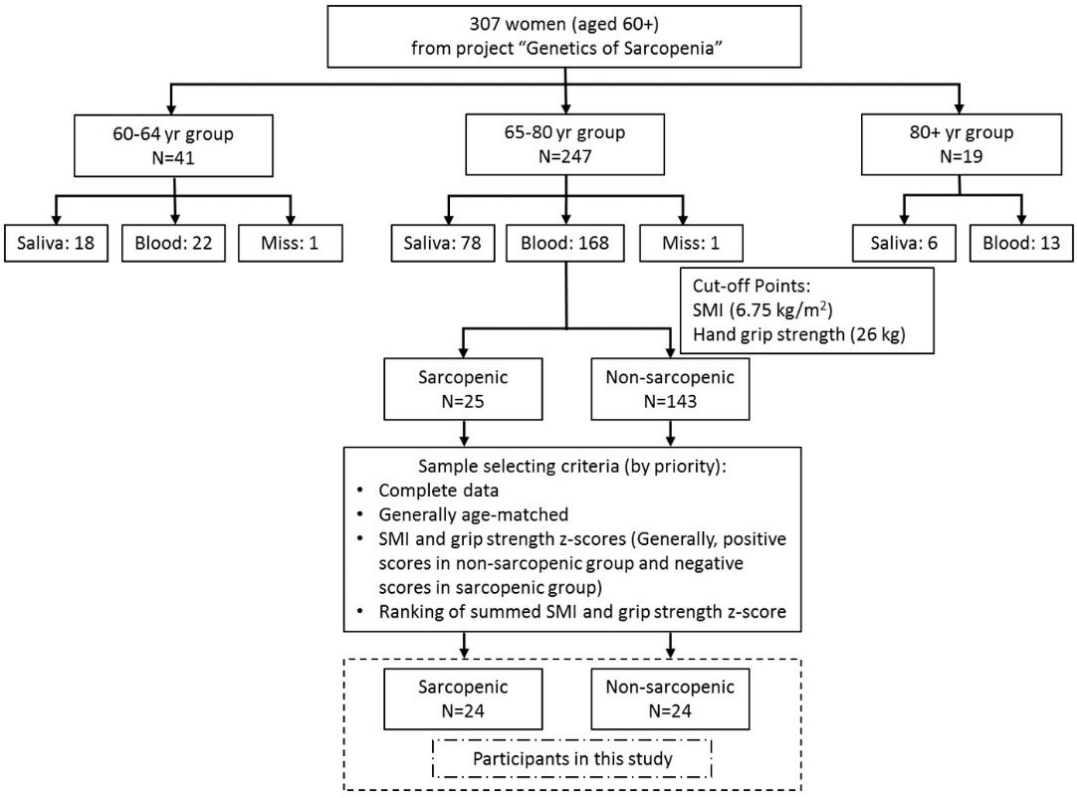


Figure 1. Flowchart of participants in the present study.

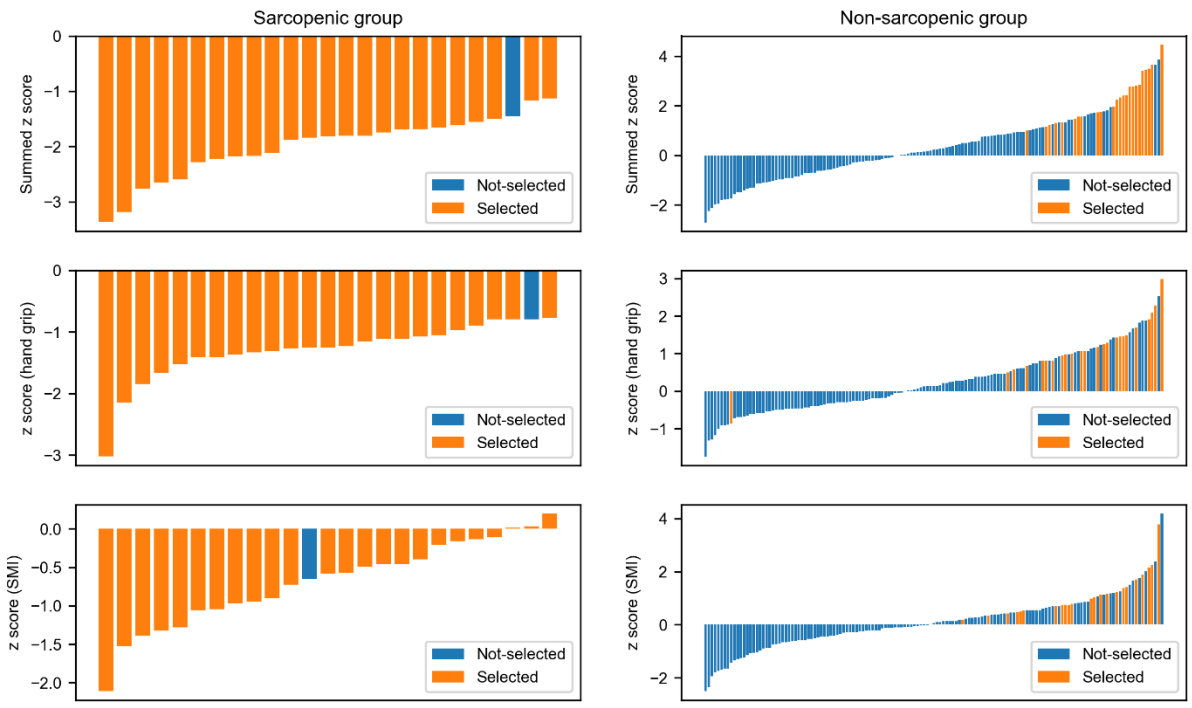


Figure 2. Distribution of Z scores in participant screening for DNA methylation analysis.

2.2. Hand grip and skeletal muscle mass measurement

HGS was measured by digital handgrip dynamometer (Jamar Plus+, JLW Instruments, Chicago, IL, US). Participants were asked to stand straight and to keep their testing arms straight out during the measurement. Verbal encouragement was given and three attempts were made on both hands. The highest value was kept for further analysis.

Electrical resistance of the body was measured by bioelectrical impedance analysis (BIA) (Bodystat 1500MDD, Bodystat Ltd, Douglas, UK). Before the test, participants were asked to remove any metal attachments and to lay in a supine position on a physiotherapy bed for 4 min. Electrodes were placed on the dorsum of the right hand and right foot according to manufacturer instructions. During the test, the participant was asked to stay quiet and relaxed. Skeletal muscle mass was estimated using the following equation which was developed by Janssen et al.[140]: Skeletal muscle mass (kg) = ($Ht^2/R \times 0.401$) – age × 0.071 + 5.102 where Ht is height in cm; R is BIA resistance in ohms; age is in years. This equation has a high coefficient of determination (r^2 = 0.86) and low bias (SEE = 2.7 kg) compared to MRI for skeletal muscle mass estimation across an age range of 18–86 yr.[140] Whole body SMI was later calculated by dividing skeletal muscle mass by height squared.

2.3. DNA extraction and methylation measurement

DNA was extracted from venous blood samples by QIAamp® DNA Blood Mini Kit (Qiagen, Crawley, UK) following the instructions of the manual. DNA methylation was measured using Illumina® Infinium MethylationEPIC BeadChip arrays (Illumina Inc., San Diego, CA, US) at the Genomics Core facility (Center for Human Genetics – UZ/KU Leuven – Herestraat 49 bus 602, B-3000 Leuven). Methylation files were read by R "Minfi" package,[141] background signals were corrected by normal-exponential out-of-band (Noob) method, and methylation values (β values, methylation percentages at measured probes) were normalized for blood cell composition by R 'FlowSorted.Blood.EPIC' package.[142] Probes were dropped under one of the three conditions: 1) probes with non-significant background signal levels ($p > 0.01$) at methylated and unmethylated channels; 2) probes that contain either single nucleotide polymorphisms at the CpG interrogation or at the single nucleotide extension as suggested in the "Minfi" package (reference array: "IlluminaHumanMethylationEPIC", annotated by ilm10b4.hg19); 3) cross-active probes that were reported in the first supplementary table of Pidsley's study.[143] A final total of 788,074 probes were kept for further methylation analyses using Partek Genomics Suite V.7.0 (Partek Inc., St. Louis, MO, US), in which CpG probes were annotated based on "HumanMethylation850" reference, "MethylationEPIC_v-1-0_B4" annotation file, "Homo sapiens" species and hg19 genome build. Notably, the DNA methylation analysis in Partek was based on the M value (log transformed methylation-to-unmethylation ratio at each CpG site, the default setting of the software) instead of the β value.

2.4. Statistics

T-tests were used to compare descriptive data (age, height, body mass, BMI, SMI and HGS) and methylation levels between the sarcopenic and non-sarcopenic groups with a significance of 0.05. Benjamini Hochberg method[236] was used for *p* value adjustment in the identification of dmCpG sites and pathway analysis. Since the *p* value of the methylation value comparison at each CpG site was greater than 0.05 after Benjamini Hochberg correction, an unadjusted *p* value threshold of 0.01 was used to define dmCpG sites. The average methylation level of analysed CpG sites, proportions of hypermethylated (defined as higher M values in the sarcopenic group than the non-sarcopenic group) and hypomethylated (defined as lower M values in the sarcopenic group) dmCpG sites, and significant genes or dmCpGs that

have been identified in previous ageing-related or muscle-related methylation studies were compared with the results of our study. Gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses (databases till May 2019) were also conducted and compared between studies based on annotated “gene symbols” of dmCpG sites using Partek.

3. Results

3.1. Descriptive data of the sarcopenic and non-sarcopenic groups

Consistent with the classification criteria, participants in the non-sarcopenic group had significantly higher SMI ($p < 0.001$) and HGS ($p < 0.001$) than that in the sarcopenic group while there was no significant age difference between the two groups ($p = 0.070$). Moreover, body mass ($p = 0.003$) and BMI ($p = 0.006$) were also significantly larger in the non-sarcopenic group in comparison with the sarcopenic group (Table 1).

Table 1. Descriptive data of participants by groups

Group	Age (year)	Body mass (kg)	Height (m)	BMI (kg/m^2)	SMI (kg/m^2)	HGS (kg)
Non-sarcopenic	70.5 ± 3.3	71.7 ± 12.8	1.60 ± 0.05	27.9 ± 4.9	7.45 ± 0.67	36.0 ± 3.7
Sarcopenic	72.5 ± 4.2	61.5 ± 9.4	1.56 ± 0.11	24.4 ± 3.4	6.00 ± 0.47	23.2 ± 2.5
p value	0.070	0.003*	0.154	0.006*	<0.001*	<0.001*

*: significant difference between the sarcopenic and the non-sarcopenic groups

3.2. DNA methylation levels

We compared methylation values at each of the analysed 788,074 CpG sites between the sarcopenic and non-sarcopenic groups but no significant CpG sites were found after Benjamini Hochberg false discovery rate (FDR) control at a level of 0.05. Therefore, CpG sites with unadjusted *p* values < 0.01 were identified as dmCpG sites (n = 6,258) (P3-Supplementary Table 1A, Figure 3A).

In our study, the total methylation level, represented by the mean methylation value of all analysed CpG sites, showed no significant difference between the sarcopenic and non-sarcopenic groups ($p = 0.812$, P3-Supplementary Table 1B, Figure 3B), while the average methylation value of dmCpGs was significantly lower in the sarcopenic group ($p = 0.004$, P3-Supplementary Table 1B, Figure 3C). When comparing average methylation values of dmCpGs by gene regions, the sarcopenic

group had significantly higher methylation levels in gene promoters (TSS200) and lower methylation levels in gene body and 3'UTR regions (P3-Supplementary Table 1B, Figure 4). Moreover, in the sarcopenic group, a greater level of methylation at CpG islands was not only found in promoter regions, but also in some intragenic regions, such as Exon 1, 3'UTR and 5'UTR (P3-Supplementary Table 1C).

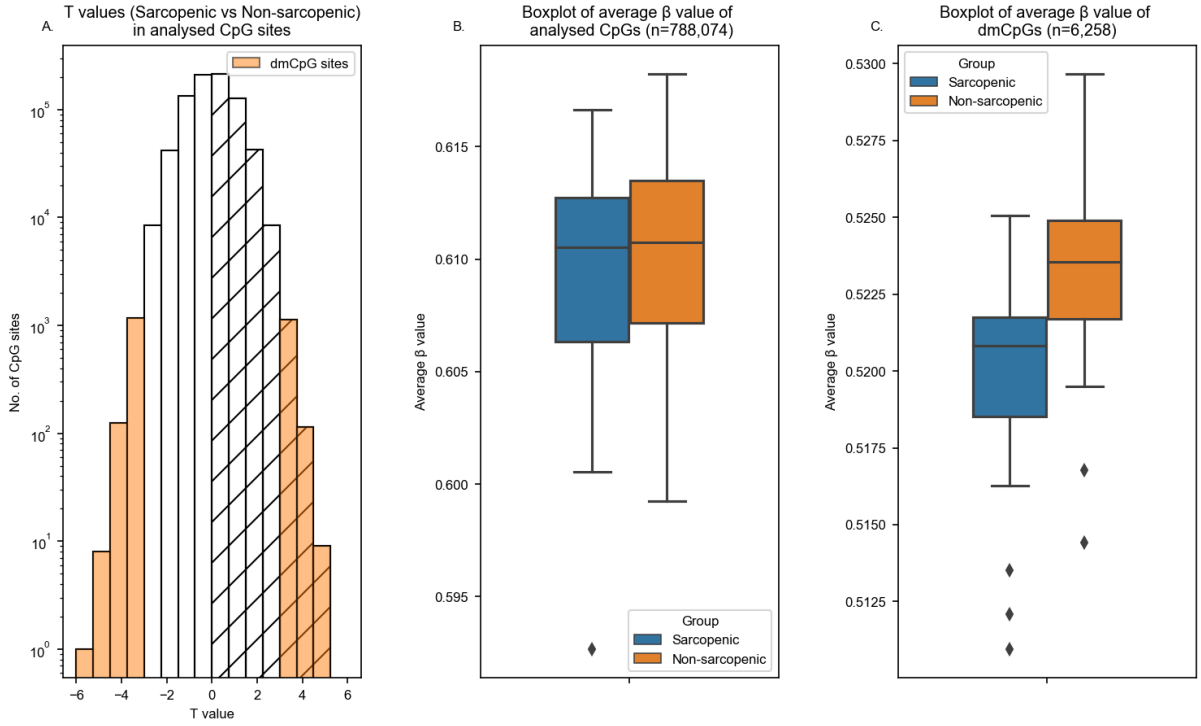


Figure 3. Distribution of T values and β values in analysed CpGs and dmCpGs. (A) Distribution of t values of analysed CpGs and dmCpGs, Areas with forward slash (/) highlight CpGs with positive T values. Yellow areas marked out dmCpGs. (B) Boxplot of average β values of the CpGs between sarcopenic and non-sarcopenic women (p = 0.812), (C) Boxplot of average β values of dmCpGs between sarcopenic and non-sarcopenic women. The β values in the sarcopenic group are significantly lower than the non-sarcopenic group (p = 0.004).

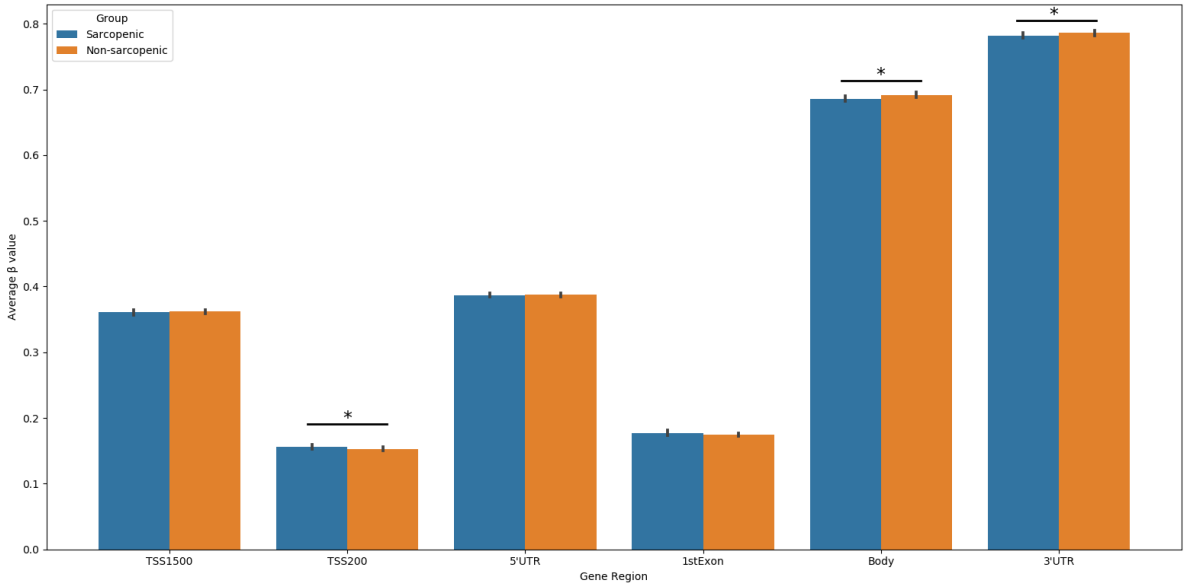


Figure 4. Comparison of average β values in different gene regions between sarcopenic and non-sarcopenic women. The sarcopenic group has significantly higher methylation levels in TSS200 (p = 0.009) and lower methylation levels in gene body and 3'UTR regions (p < 0.001 and p = 0.021, respectively).

Among those identified dmCpG sites, 51.2% (n = 3,205) were hypermethylated and the remaining 48.8% (n = 3,053) were hypomethylated (P3-Supplementary Table 1D, Figure 5). dmCpG methylation value-based unsupervised clustering showed that the majority of participants were clustered by corresponding groups (Figure 6).

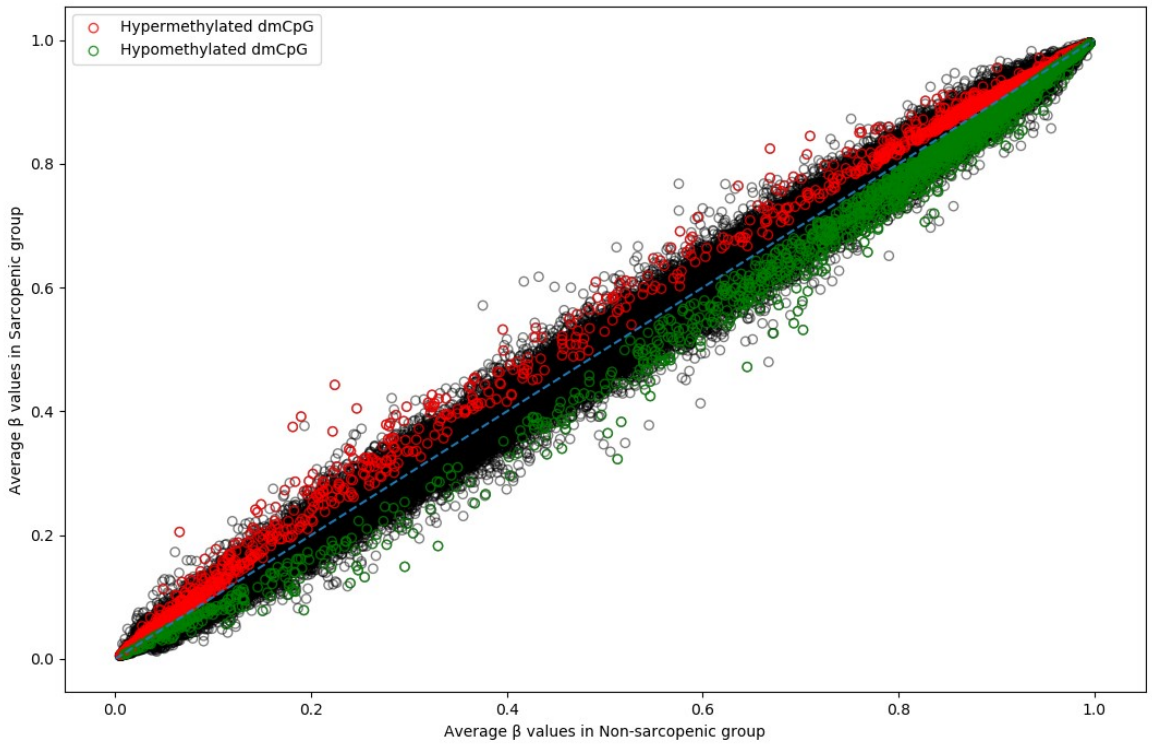


Figure 5. Scatter plot of average β values of CpGs in sarcopenic versus non-sarcopenic women with hypermethylated dmCpGs in red and hypomethylated dmCpGs in green.

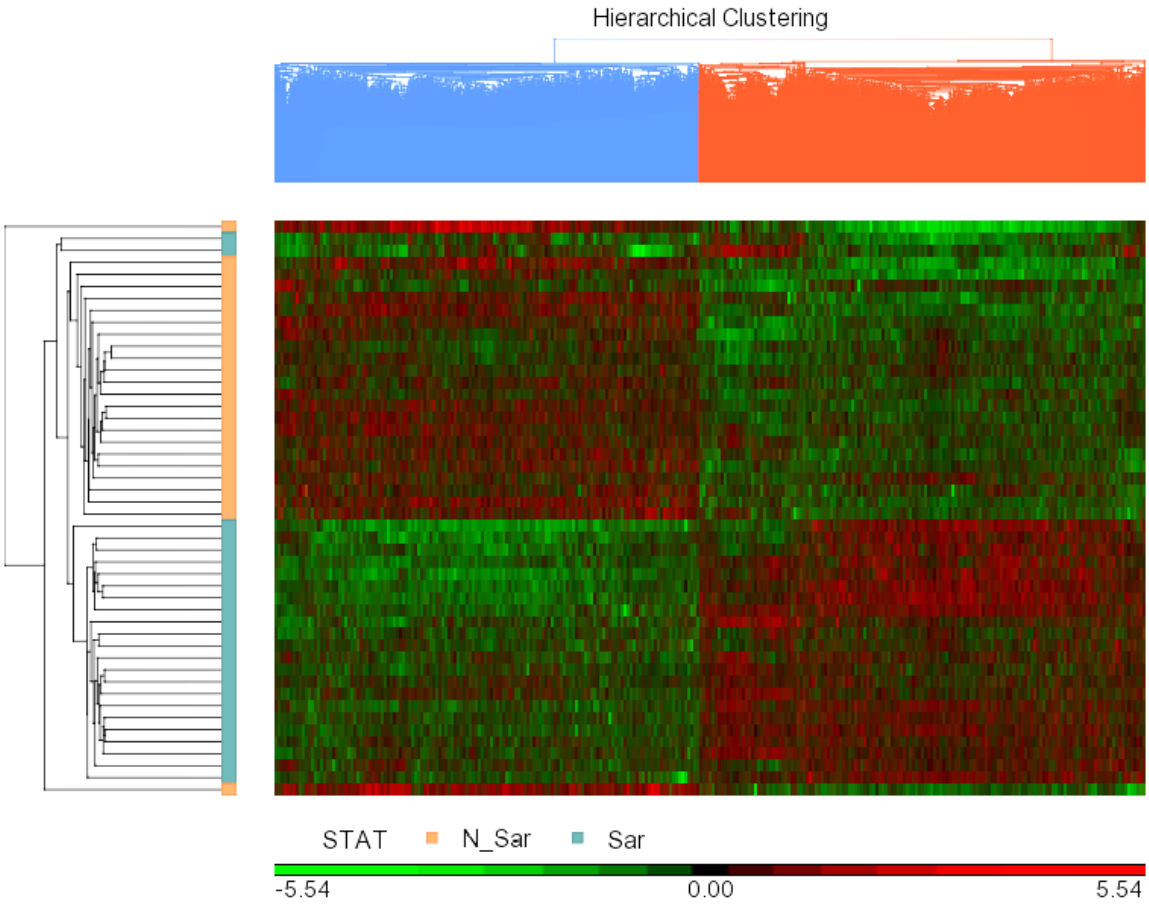


Figure 6. Hierarchical clustering of dmCpGs. The left bar represents each participant (n=48). The hierarchical cluster on the top represents the two clusters based on methylation status of dmCpGs (green colour stands for a negative M value, red colour stands for a positive M value). The majority of participants were clustered into two groups based on the methylation levels of dmCpGs

3.3. Distribution of differentially methylated CpG sites

Although a large amount of dmCpGs were located in CpG-poor areas (known as the Open Sea), CpG islands had the highest methylation rate among all the analysed CpG regions (1.19%, Figure 7, P3-Supplementary Table 1D). dmCpGs distribution included 28.6% located in CpG islands, 9.8-10.6% (totalling 20.4%) in CpG shores (within 2kb of CpG islands[225]) and 2.5-3.1% (totalling 5.6%) in CpG shelves (within 2kb of CpG shores[225]) (Figure 8). This indicated that the proportion of dmCpGs was negatively related to the distance away from the CpG island. Meanwhile, 83.8% of dmCpG sites located in CpG islands were hypermethylated while CpG south shelf

(S_Shelf) had the largest hypomethylated proportion of 80% (Figure 9).

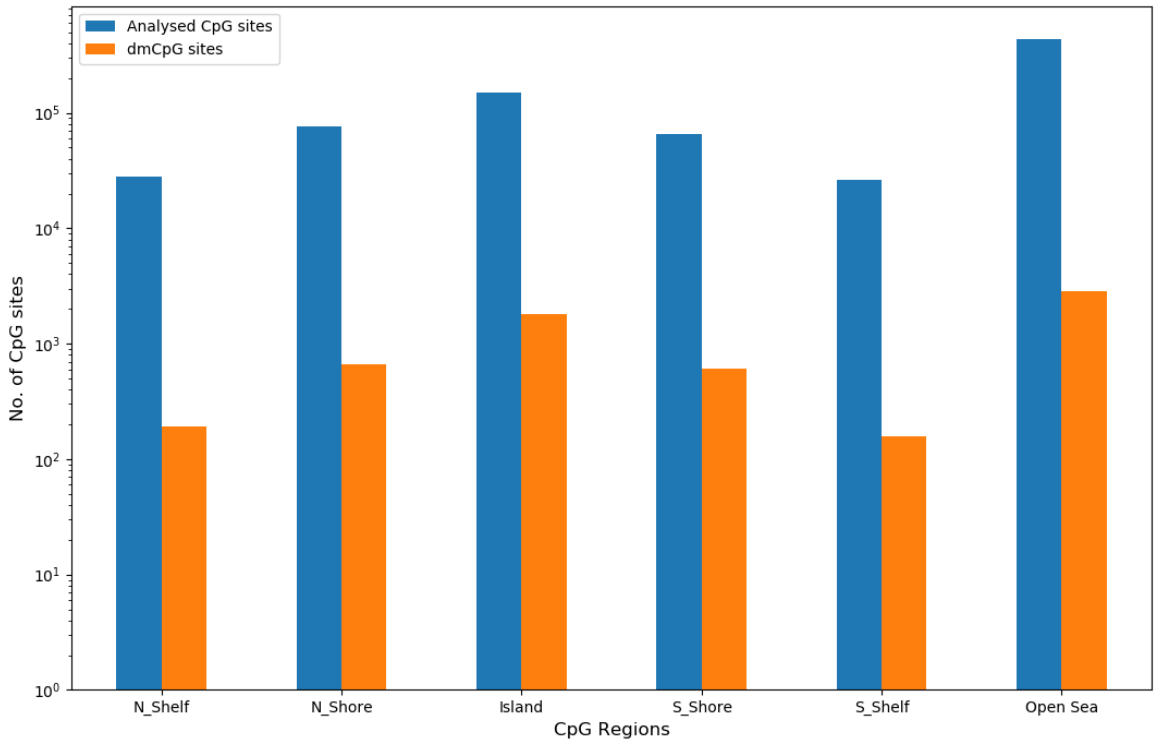


Figure 7. Distribution of analysed CpGs and dmCpGs by CpG regions. Most of the identified dmCpGs located in the Open Sea region (probably due to the high proportion of analysed CpGs in this region). The CpG island region contributed the second most dmCpGs.

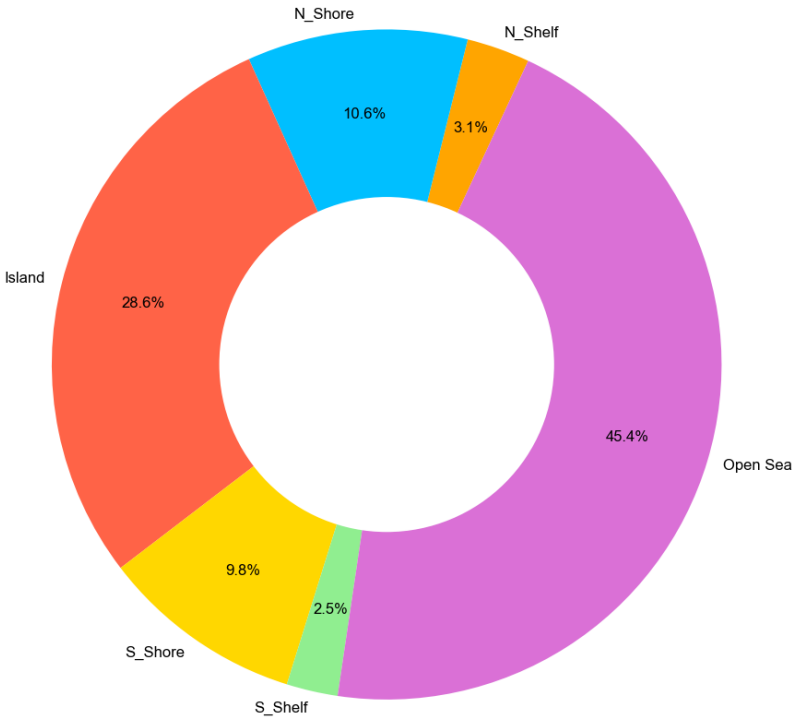


Figure 8. dmCpGs distribution includes 28.6% located in CpG islands, 9.8 in CpG

south shores, 10.6% in CpG north shores, 2.5% in CpG south shelves and 3.1% in CpG north shelves. This indicates that the proportion of dmCpGs is negatively related to the distance away from the CpG island.

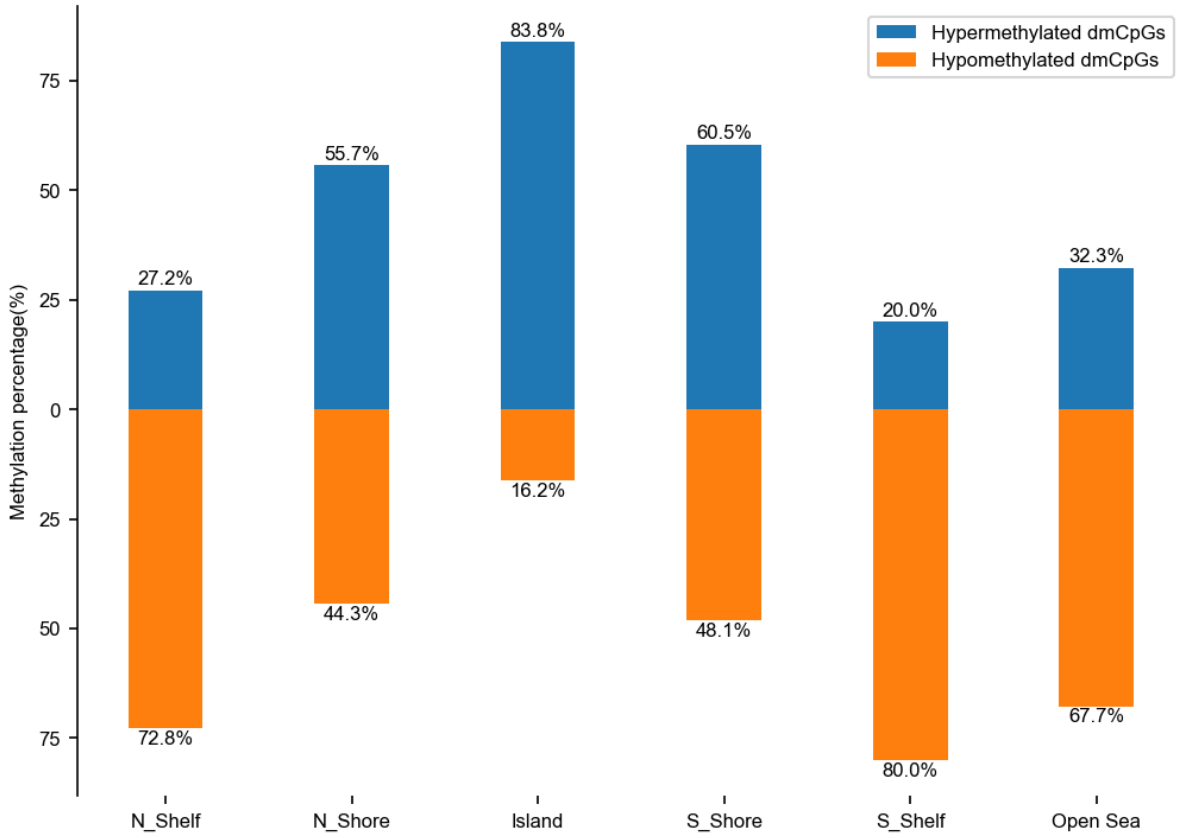


Figure 9. Methylation status of dmCpGs in CpG regions. dmCpGs with higher M values in the sarcopenic group than the non-sarcopenic group are defined as hypermethylated. dmCpGs with lower M values in the sarcopenic group are defined as hypomethylated.

Besides analyses on CpG regions, we also analysed the distribution of dmCpGs by chromosomes. The largest amount of dmCpG sites were found in chromosome 1 (Figure 10A) while chromosome 19 and chromosome 18 had the largest (1.0%) and the smallest proportion (0.6%) of dmCpGs in analysed CpG sites, respectively (P3-Supplementary Table 1D, Figure 10B). Furthermore, chromosome 19 had the highest percentage (1.1%) of hypermethylated dmCpG sites in analysed hypermethylated sites while chromosome X had the highest hypomethylated proportion (1.2%), followed by chromosome 16 (1.0%).

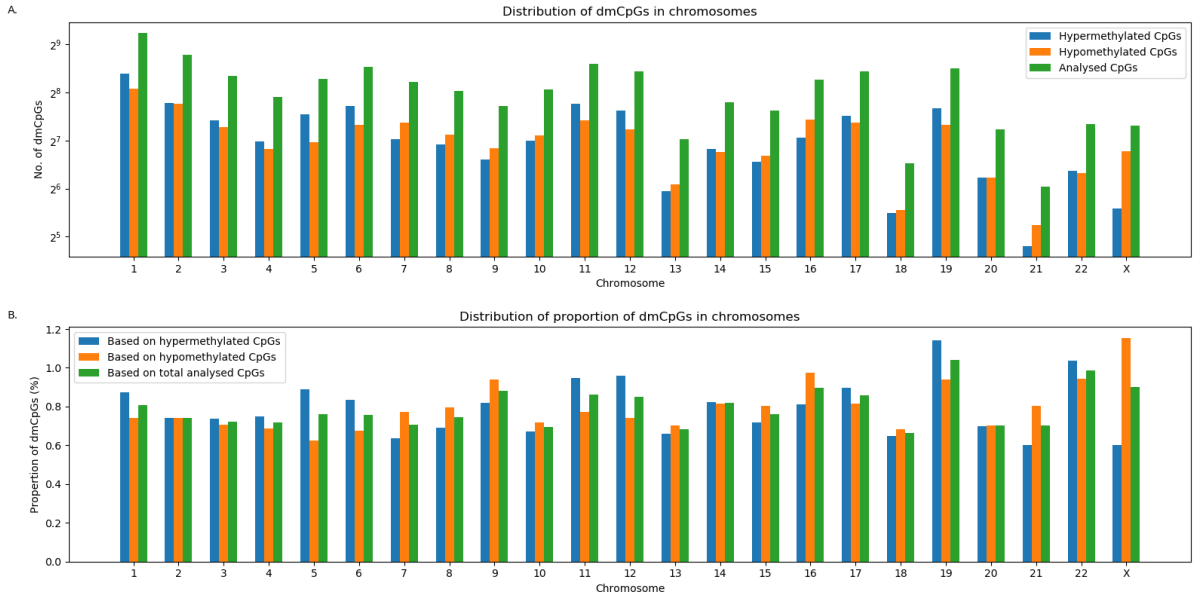


Figure 10. Distribution of dmCpGs and dmCpG proportions across chromosomes. (A) Distribution of dmCpGs over chromosomes, (B) Chromosomal distribution of relative dmCpGs proportions. Chromosome 19 has the largest relative dmCpG methylation proportion (1.0%) while chromosome 18 has the smallest relative dmCpG methylation proportion (0.6%).

3.4. Genes and gene regions with altered methylation status

Among the identified 6,258 dmCpG sites, 4,840 dmCpGs were annotated by gene names and regions (based on the "UCSC_RefGene_Name" and "UCSC_RefGene_Group" columns in P3-Supplementary Table 1A). In total, hypermethylation and hypomethylation were found in 2,422 and 1,913 genes, respectively (some CpG sites are annotated with multiple gene names and regions, P3-Supplementary Table 2A). With the largest number of analysed CpG sites (n = 7,572, P3-Supplementary Table 2B) among all annotated genes, *PC* gene contributed the largest amount of dmCpGs (n = 71, P3-Supplementary Table 2B), which were located in hypermethylated CpG islands of the gene body region (P3-Supplementary Table 2M). Meanwhile, 279 genes were identified with both hyper- and hypomethylation (P3-Supplementary Table 2B, Figure 11). Hypermethylation was more common than hypomethylation in promoter regions (TSS1500 and TSS200), 5'UTR and Exon 1 (P3-Supplementary Table 2, Figure 12). Notably, we found that the methylation status of some genes was not identical across the same gene region. From our results, gene promoters of 16 genes and gene bodies of 97 genes were found with both hyper- and hypomethylation (P3-Supplementary Table 2D).

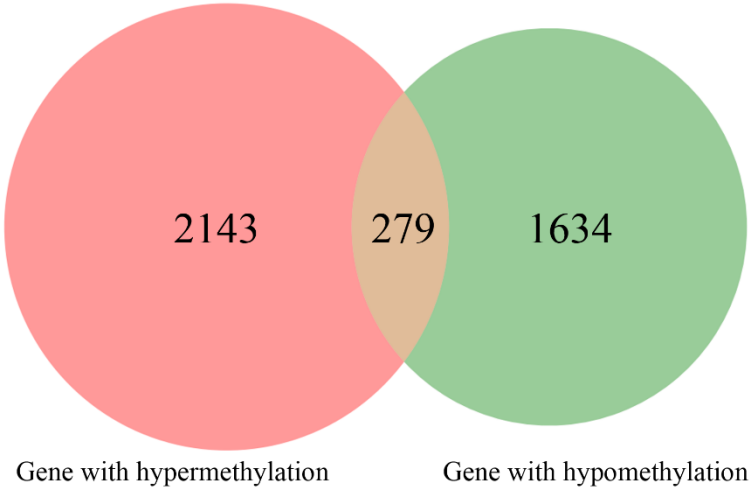


Figure 11. Venn diagram of genes annotated by dmCpGs. 2,143 genes are found only with hypermethylated dmCpGs and 1,634 genes are found only with hypomethylated dmGpGs. There are 279 genes with both hyper- and hypomethylated dmCpGs.

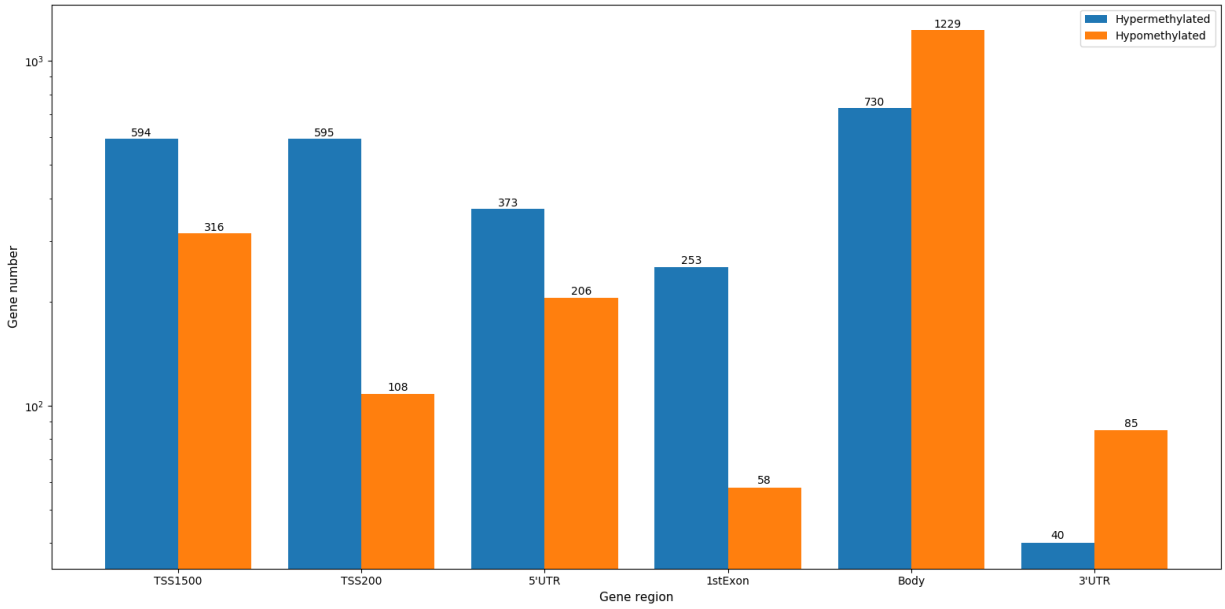


Figure 12. Distribution of methylated genes by gene regions. The majority of dmCpGs located in the gene body, the promoter region contributed the second most dmCpGs.

3.5. Enrichment and pathway analysis

633 terms in GO enrichment analysis were identified with significance after FDR control (q value < 0.05) based on the “gene symbols” of dmCpG sites (P3-Supplementary Table 3A). The most significant GO term was “protein binding” which included 1,680 hypermethylated CpG sites (1,488 genes) and 1,356

hypomethylated CpG sites (1,152 genes) (P3-Supplementary Table 3C, E). Most hypermethylated CpGs located in the CpG island (P3-Supplementary Table 3D) while most hypomethylated CpGs were found in the Open Sea (P3-Supplementary Table 3F). There were 197 genes with both hyper- and hypomethylated CpGs in the "protein binding" term (P3-Supplementary Table 3G). GO analysis based on hyper- and hypomethylated promoter regions (P3-Supplementary Table 1E, F) showed that 291 GO terms were significantly related to hypermethylated promoter regions (q value < 0.05) with the term "intracellular part" as the most significant term while 10 GO terms were significantly connected to hypomethylated promoter regions (q value < 0.05) with the term "androgen receptor binding" as the most significant term (P3-Supplementary Table 3H, I).

No significant KEGG pathways (q value < 0.05) were found after FDR control while 37 pathways showed unadjusted significance (*p* value < 0.05) (P3-Supplementary Table 4A). These unadjusted significant terms covered many muscle-related aspects such as muscle function (e.g. apelin signaling, cGMP-PKG signaling, insulin resistance), actin cytoskeleton regulation (e.g. phosphatidylinositol signaling, focal adhesion, adherens junction), energy metabolism (e.g. thermogenesis, AMPK signaling, glucagon signaling), neural control (e.g. axon guidance, GABAergic synapse), signal transduction (e.g. Wnt signaling, MAPK signaling, cAMP signaling), blood pressure regulation (e.g. aldosterone) and cell regeneration (e.g. cell cycle, oxytocin signaling). As the most significant KEGG pathway, the "apelin signaling pathway" included 56 dmCpGs, half of which were hypermethylated (P3-Supplementary Table 4C, E, J). Most hypermethylated CpGs located in the CpG island (P3-Supplementary Table 4D) and the majority of hypomethylated CpGs were found in the Open Sea (P3-Supplementary Table 4F). There were five genes with both hyper- and hypomethylated CpGs in the "apelin signaling pathway" (P3-Supplementary Table 4G). KEGG analysis based on hyper- and hypomethylated promoter regions showed that the "cell cycle" and "thermogenesis" pathways were the most significant terms associated with hyper- and hypomethylated promoter regions, respectively (P3-Supplementary Table 4H, I).

3.6. Comparative analysis with previous studies

We first compared muscular phenotype-related genes, which have been reported in previous studies,[66,212,237] with the genes identified in our study (based on the "UCSC_RefGene_Name" column, P3-Supplementary Table 1A) and found 34

genes in common (P3-Supplementary Table 5A). We further compared our dmCpGs, annotated genes and significant KEGG pathways (unadjusted, *p* value < 0.05) with those previously reported in DNA methylation studies on ageing,[101,229,230] muscle development,[238] muscle diseases[234,235] and exercise.[102,104,105,195,228]

When compared with studies of blood sample methylation during ageing,[229,230] we found seven common CpGs in total and the CpGs showed consistent methylation direction in all studies (P3-Supplementary Table 5B). Zykovich et al.[101] studied the skeletal muscle methylation difference between older and young participants and identified 5,963 ageing-related dmCpGs, among which 35 dmCpGs were found in our dmCpGs and 11 CpGs had the same methylation direction in both studies (P3-Supplementary Table 5B). Two dmCpGs (cg10093679 and cg19291355) in our study were found in the top 500 ageing-related dmCpGs suggested by Zykovich et al. and the methylation status was identical in both studies (P3-Supplementary Table 5B). Moreover, Zykovich et al. located 17 ageing-related intragenic dmCpGs (with 16 CpGs being hypermethylated) in *NFATC1*, a gene closely associated with muscle function as it codes for a transcription factor promoting the expression of slow fiber types and is involved in neuromuscular signal conduction. In our results, three dmCpGs were located in intragenic regions of *NFATC1* and all were hypomethylated (P3-Supplementary Table 5B). However, none of these CpGs were included as dmCpGs in Zykovich's study.

Four CpGs from three genes (*HOXD4*, *SEPT9* and *MBP*) (P3-Supplementary Table 5C) were differentially methylated in both our study and muscle inflammatory disease in children.[235] By comparing with the methylation study on rhabdomyosarcoma,[234] we found 360 genes in common. GO and KEGG analysis on those genes showed that many of the significant terms were associated with muscle function, DNA transcription regulation, nervous system development and signal transduction (P3-Supplementary Table 5D, E).

Seaborne et al.[102] identified several DNA methylation sites that were closely related to resistance training. When comparing our dmCpGs with the top 500 resistance training-related CpG sites in Seaborne's study, only two sites (cg00077516 and cg09739536) were found in common (P3-Supplementary Table 5F). We found 17 common CpGs when comparing with loading-related CpGs and 8 CpGs showed different fold change directions in our study and Seaborne's study (P3-Supplementary Table 5F). We also identified 9 common CpGs by comparing with

unloading-related CpGs and only one CpG had the same direction in both studies (P3-Supplementary Table 5F). Turner et al.[104] made a thorough analysis of gene expression and DNA methylation induced by acute and chronic resistance training. We shared two CpGs (cg08284143 and cg11692073) that showed hypermethylation after acute resistance training (based on Turner's study) and hypomethylation in sarcopenia (based on our study) (P3-Supplementary Table 5G). Six common CpGs were found when comparing with hypomethylated CpGs after acute resistance training and all of them had the same fold change direction in both studies (P3-Supplementary Table 5G). When comparing with dmCpGs related to chronic resistance training, we identified seven common CpGs, five of which had the same fold change direction in both studies (P3-Supplementary Table 5G). We also found 19 genes (24 CpGs) out of 51 genes that had significant epigenetic and transcriptome changes under all acute/chronic training/detraining conditions reported in Turner's study (P3-Supplementary Table 5G). Furthermore, the *SRGAP1*, *PLXNA2* and *JOSD1* were found with increased gene expression after resistance training (based on Turner's study) while hypermethylated gene promoters were found in sarcopenia (based on our study) (P3-Supplementary Table 5H). We compared unadjusted significant KEGG pathways (p value < 0.05) between our study and Turner's study and found three pathways (cGMP-PKG signaling pathway, human papillomavirus infection and proteoglycans in cancer) in common (P3-Supplementary Table 6A-K). Notably, the counts of hyper- and hypomethylated CpGs in the pathway "cGMP-PKG signalling" and "human papillomavirus infection" were very similar while the amount of hypomethylated CpGs was twice as many as the hypermethylated CpGs in the pathway "proteoglycans in cancer". Sailani et al.[105] identified 748 gene promoters with significant methylation difference between physically active and inactive older participants. By comparing with our hyper- and hypomethylated gene promoters (P3-Supplementary Table 1E, F), we found 23 common genes with hypermethylated promoters and 369 genes with hypomethylated promoters (P3-Supplementary Table 6L). The significant pathways based on these common genes were associated with energy metabolism, signal transduction, myogenesis and actin cytoskeleton regulation (P3-Supplementary Table 6M, N).

4. Discussion

4.1. DNA methylation patterns

The ageing-related DNA methylation pattern is characterized by globally decreased and regionally (CpG islands and shores) increased methylation levels.[91] By comparing the methylation status of blood CD4+ cells between newborns and centenarians, Heyn et al.[229] reported a decreased global methylation level in older participants. However, the study of Zykovich et al.[101] revealed a global trend of hypermethylation in ageing skeletal muscle. In our study, we found no significant difference in total methylation level of all analysed CpGs between the sarcopenic and the non-sarcopenic groups. The even distribution of hyper- and hypomethylated dmCpGs in our study also contradicted the findings of Bell et al.[230] (whole blood-based) and Zykovich et al.[101] (skeletal muscle-based) where more than 90% of ageing-related dmCpGs were hypermethylated. Such inconsistency in DNA methylation might be partly related to the relatively narrow range of age of our participants. Since our participants in both groups came from a similar age range (65–80 years) with no difference between them, the age effect on DNA methylation might be partially controlled. Nevertheless, the methylation condition of CpG islands in gene promoters was quite similar between sarcopenia-related and ageing-related methylation patterns. In our study, promoter CpG islands of the sarcopenic group were hypermethylated when compared with those of the non-sarcopenic counterpart. Similarly, a higher methylation level was also found in promoter CpG islands of centenarians in comparison with newborns.[229]

Hypermethylated CpG islands in gene promoters have been associated with the down-regulation of genes.[223] In our study, the sarcopenic group had a higher methylation of promoter CpG islands than the non-sarcopenic group, perhaps indicating that reduced function of corresponding genes accompanies sarcopenia. However, gene expression was not studied in the present manuscript and therefore would require further analysis to confirm this assumption. Unlike promoters, significant hypomethylation was found in gene bodies and 3'UTRs of the sarcopenic group. Gene bodies in the human genome are prevalently methylated.[239] Yet, the connection between methylated gene bodies and gene expressions remains debatable. Mendizabal et al.[240] studied nearly 2,000 genes, the gene bodies of which were commonly hypomethylated in normal tissues, and found that those genes were prone to have significantly hypomethylated gene bodies in cancer

samples. Such results suggested that hypomethylated gene bodies were more related to cancer-associated dysregulation, which is supported by Yang et al.,[241] who showed that demethylated gene bodies could cause down-regulation in gene expression. Contradictorily, by analysing cell-lines, Jjingo et al.[242] reported a bell-shaped relationship between gene transcriptions and methylated gene bodies. Therefore, further measurements on gene expression are needed to evaluate the effect of methylated gene bodies on the activity of corresponding genes. Furthermore, although most of the genes identified in our study showed a single methylation pattern, we found some genes with both hyper- and hypomethylation in the same gene region, indicating that the sarcopenic condition might be related to the interaction between methylation and demethylation in these genes.

4.2. dmCpG-related GO terms and KEGG pathways

The “protein binding” was the most significant GO term identified in our study (P3-Supplementary Table 3A), indicating that signal transduction and cellular metabolism were closely related to our dmCpGs. The “apelin signaling pathway” was the mostly enriched pathway in KEGG analysis (P3-Supplementary Table 4A). Apelin is a peptide that can reduce arterial stiffness,[243] and enhance muscle mitochondriogenesis[244] and protein synthesis.[245] The endogenous apelin level decreases with age while apelin treatment can induce muscle mass and reverse age-associated sarcopenia in mice.[245] Moreover, apelin can be up-regulated by insulin and the increased concentration of plasma apelin has been found in obese humans and mice.[246] Therefore, our finding of the “apelin signaling pathway” not only suggests a close association between apelin and muscle degeneration but also indicates a possibility of sarcopenia-related alterations in energy metabolism and body composition that might result in sarcopenic obesity.[6]

Since hypermethylated gene promoters are related to repression of gene expression,[223] we conducted GO analysis based on genes with hyper-/hypomethylation in promoter regions to explore possible biological processes that might be different between the sarcopenic and non-sarcopenic groups. From GO results, many metabolism-related processes and cellular components were significantly related to genes with hypermethylated promoter regions (P3-Supplementary Table 3H), indicating that cellular metabolism might be downregulated in the sarcopenic group when compared with the non-sarcopenic group. Meanwhile, based on genes with hypomethylation in promoter regions,

several terms related to negative regulations of the insulin signaling pathway have been identified (P3-Supplementary Table 3I). Such findings suggest that negative regulation of insulin signaling might be strengthened in the sarcopenic group, possibly resulting in restricted glucose uptake in skeletal muscles[247] and reduced protein synthesis via downregulated activations of insulin receptor and PI3K.[248] Moreover, the associations between hormone receptor bindings terms (e.g. androgen and steroid hormone) and hypomethylated promoters (P3-Supplementary Table 3I) also indicate an overexperssion of hormone receptors in sarcopenia to compensate decreased plasma hormone levels that are important in maintaining muscle and bone mass.[249] However, hormone measurements are required for further confirmation.

The “cell cycle” and “thermogenesis” pathways identified by genes with hyper- and hypomethylated promoter regions (P3-Supplementary Table 4H, I) suggest possibly decreased cell regeneration ability and increased heat production associated with sarcopenia. Since the sarcopenic group has lower body mass and BMI than the non-sarcopenic group, the sarcopenic group might have less fat for heat preservation, resulting in elevated heat production for body temperature maintenance. There is also a possibility that the sarcopenic group has less body fat because of increased heat production.

4.3. Gene, dmCpG and pathway comparisons with previous studies

The current study highlights several muscle-related genes which have been identified in previous muscle mass- or muscle strength-related studies,[66,212,237] among which is the gene *VDR* (P3-Supplementary Table 5A). *VDR* encodes vitamin D receptor, a protein widely known for regulation of calcium and phosphate homeostasis.[250] *VDR* gene knockout mice have impaired bone and mineral metabolism and rickets after weaning.[251] Many studies of ageing have also connected decreased *VDR* expression to osteoporosis,[252] frailty,[253] low muscle strength[169] and low muscle mass.[212] In our study, one hypermethylated CpG island was indeed in the *VDR* promoter region. This indicates a possible reduced expression of *VDR* in the sarcopenic group and could partially explain the decline in muscle function and increased risk of frailty associated with sarcopenia.[6]

We also identified 42 genes (P3-Supplementary Table 5B) that were reported in previous ageing-related methylation studies.[101,229,230] Many of these genes are also closely muscle-related. For example, we found five genes (*HSPB1*, *PBX4*, *FZD5*,

HMGA1 and *AARS2*) with sarcopenia-related hypermethylation in CpG islands (P3-Supplementary Table 5B). *HSPB1* encodes heat shock protein beta-1, which is a member of the small heat shock protein family that plays an important role in muscle development, differentiation and protection against heat and mechanical stress.[254] Mutations of *HSPB1* have been related to distal hereditary motor neuropathy, muscle weakness and fat infiltration.[255] The protein Pre-B-cell leukemia transcription factor 4 (encoded by *PBX4*) is a member of the Pbx family, which facilitates binding of MyoD to gene regulatory regions to initiate skeletal muscle differentiation.[256] In zebrafish, Pbx helps regulate the development of fast-twitch skeletal muscle[257] that undergoes preferential atrophy during ageing in humans.[258] *AARS2* encodes mitochondrial alanyl-tRNA ligase, a member of the aminoacyl-tRNA synthetase family that plays an important role in mRNA translation. Mutations of *AARS2* have been related to neurodegenerations[259] and cardiomyopathy.[260] The hypermethylated CpG island of the *AARS2* promoter region in our study suggests a down-regulation of this gene's activity in sarcopenia. *FZD5* encodes receptors for the Wnt5A ligand, which is connected with muscle mass regulation via the mTOR pathway.[261] *HMGA1* is involved in multiple cellular processes such as DNA repair, transcriptional regulation and cell cycle regulation. Overexpression of *HMGA1* has been found in cancer, indicating the association of *HMGA1* with cell regeneration.[262] The *SRGAP1* promoter region was hypomethylated after resistance training[104] and a hypermethylated promoter region was found in our study (P3-Supplementary Table 5H). *SRGAP1* encodes a GTPase activator that regulates cell regeneration and axon guidance and knockdown of *SRGAP1* suppresses cell proliferation by inhibiting the Wnt/β-catenin pathway.[263] These comparisons indicate a possible connection between *SRGAP1* and exercise-induced muscle growth as well as muscle degeneration during ageing.

Although the pattern of DNA methylation is mostly tissue-specific,[264] there are still some overlaps between our dmCpGs and those that have been previously reported in muscle samples. In a recent study on muscle DNA methylation changes induced by resistance training and detraining, Seaborne et al.[102] reported a significantly decreased methylation in dmCpG site cg09739536 from gene *ZFP2* after a 7 weeks of resistance training. This finding is consistent with our observation that the same CpG site was significantly hypermethylated in the sarcopenic group and that less methylation of *ZFP2* is associated with greater skeletal muscle mass and strength. Similarly, six genes (*SNHG5*, *FUCA2*, *MUC5B*, *SLC30A1*, *CCR3*, and *SMEK2*) were

found in the loading phase of Seaborne's study as having different fold change directions from our study, and the genes *SNHG5*, *FUCA2* and *CCR3* had methylation changes in promoter regions (P3-Supplementary Table 5F), suggesting the changes in expression of these three genes might be associated with both muscle degeneration and regeneration. Despite some similarity in methylated genes and CpG sites which we shared with previous muscle biopsy-based studies,[101,102,104,105,234,235] the methylation patterns were not always consistent. For instance, the dmCpG site cg00077516 from gene *MRPS27* was hypomethylated during detraining-induced muscle changes in one study[102] but hypermethylated in the sarcopenic group in our study. A possible explanation for this discrepancy might be the difference in physical condition of participants. Our study was based on older women susceptible to age-related muscle degeneration while the detraining phase in participants of Seaborne's study can be considered as a return to pre-training levels after a prolonged training load stimulus. It is possible that different mechanisms are involved in age-related atrophy and detraining-related muscular changes. Among the 35 common CpGs shared with the study of ageing by Zykovich et al.,[101] 24 CpGs showed a different methylation direction (P3-Supplementary Table 5B). Wang et al.[235] identified hypomethylated *HOXD4* gene in juvenile dermatomyositis patients, while we only found one hypermethylated dmCpG located in the gene body of *HOXD4*. Similar inconsistent methylation patterns were also found in comparison with the study of Mahoney et al.[234] on rhabdomyosarcoma (P3-Supplementary Table 5C). Moreover, we only identified two common CpGs from the top 500 ageing/exercise-related dmCpGs (skeletal muscle-based)[101,102] (P3-Supplementary Table 5B, F). On the other hand, all CpGs that were shared with previous blood sample-based methylation studies had the same methylation direction (P3-Supplementary Table 5B). These methylation findings confirm tissue-specific methylation differences and therefore, it is difficult to determine sarcopenia-related methylation differences in skeletal muscles of our participants based on differentially methylated genes identified from blood samples.

In respect of pathway comparative analysis, we identified three common pathways (cGMP-PKG signaling pathway, human papillomavirus infection and proteoglycans in cancer), which are associated with cell proliferation and development from previous exercise-related methylation studies[104] (P3-Supplementary Table 6E). Activation of the cGMP-PKG signaling pathway is related to the inhibition of cell proliferation.[265] Increased activation of phosphodiesterase 5, an inhibitor of the

cGMP-PKG signaling pathway, has been reported in tumor cell lines.[266] Proteoglycans play an important role in regulating muscle development. Heparan sulphate proteoglycans are essential for signal transduction in several muscle growth or differentiation pathways, such as FGF-2 and HGF.[267] The increases of biglycan and decorin found in muscle dystrophy also suggest an involvement of proteoglycans in response to myofibre damage.[268]

4.4. Limitations

Several parameters have been suggested to diagnose sarcopenia, while cut-off points vary with measurement techniques and populations.[9] Recently, cut-off points for sarcopenia tests have been advised by the European Working Group on Sarcopenia in Older People (EWGSOP).[6] In our study, however, although the assessment methods were the same (SMI and HGS), none of these recommended values were used. We did not use the advised HGS because the value given by EWGSOP was derived from the study of Dodds et al.,[269] who summarized the grip strength across twelve British studies that used different brands of dynamometers and test protocols from our study. Since we also did not have a healthy young group as a reference,[9] we used the lower quintile of HGS from our participants as one of the cut-off points to define sarcopenia, an approach having been previously used in several studies to classify sarcopenia.[270,271] Moreover, EWGSOP only provided values for appendicular SMI and not whole body SMI.[6] Considering that our SMI was calculated using whole body skeletal muscle mass estimated by BIA, to identify a second cut-off point to define sarcopenia we followed Janssen et al.,[16] who previously defined sarcopenia cut-off points using estimated whole body skeletal muscle mass in a larger sample of comparably aged participants. Besides the definition of sarcopenia, we also acknowledge that our study involves a limited sample size and the methylation value at each CpG site was compared without correction for covariates such as age and physical activity. Moreover, our assumptions on gene function changes are based on methylation status; gene expression was not studied. Therefore, further studies with a larger sample size, adjusted methylation values and gene transcriptome analysis might be more informative.

5. Conclusions

In the present study, we compared blood DNA methylation patterns between age-matched sarcopenic and non-sarcopenic older women. We identified 6,258

differentially methylated CpG sites and found that the sarcopenic group had significantly less total methylation at these sites. Pathway analyses showed that these sarcopenia-related CpG sites are involved in many muscle-related aspects, such as muscle differentiation, muscle function and energy metabolism. Moreover, sarcopenia-related hypermethylation was typically found in gene promoters and hypomethylation found more often in gene body and 3'UTR regions. Our study has therefore enriched the understanding of DNA methylation differences associated with sarcopenia.

6. Conflict of interest

The authors declare no conflicts of interest. The authors certify that they comply with the ethical guidelines for publishing in the Journal of Cachexia, Sarcopenia and Muscle: update 2017.[272]

7. Acknowledgements

This study is supported by the Research Foundation – Flanders (FWO) [grant G.0898.15], and a PhD grant to Lingxiao He and Praval Khanal by the Erasmus Mundus ‘Move-AGE’ joint doctorate program.

CHAPTER 3

Role of genetic and methylation profiles in ageing muscle phenotypes

Paper 4: Associations of combined genetic and epigenetic scores with muscle size and muscle strength: a pilot study in older women.

Paper 4

Associations of combined genetic and epigenetic scores with muscle size and muscle strength: a pilot study in older women

Lingxiao He, Praval Khanal, Christopher I Morse, Alun Williams, Martine Thomis

Submitted to *Journal of Cachexia, Sarcopenia and Muscle*.

Abstract

Background:

Inter-individual variance in skeletal muscle is closely related to genetic architecture and epigenetic regulation. Studies have examined genetic and epigenetic relationships with characteristics of ageing muscle separately, while no study has combined both genetic and epigenetic profiles in ageing muscle research. The aim of this study was to evaluate the association between combined genetic and methylation scores and skeletal muscle in older women.

Methods:

48 older Caucasian women (aged 65–79 yr) were included in this study. Biceps brachii thickness and vastus lateralis anatomical cross-sectional area ($ACSA_{VL}$) were measured by ultrasonography. Maximum isometric elbow flexion (MVC_{EF}) and knee extension (MVC_{KE}) torques were measured by a customized dynamometer. The muscle-driven genetic predisposition score (GPS_{SNP}) was calculated based on seven muscle-related single nucleotide polymorphisms (SNPs). DNA methylation levels of whole blood samples were analysed using Infinium MethylationEPIC BeadChip arrays. The DNA methylation score was calculated as a weighted sum of methylation levels of sarcopenia-driven CpG sites (MS_{SAR}) or an overall gene-wise methylation score (MS_{SNP}, the mean methylation level of CpG sites located in muscle-related genes). Linear regression models were built to study genetic and epigenetic associations with muscle size and strength. Three models were built with both genetic and methylation scores: (1) MS_{SAR} + GPS_{SNP}, (2) MS_{SNP} + GPS_{SNP}, (3) gene-wise combined scores which were calculated as the ratio of the SNP score to the mean methylation level of promoters in the corresponding gene. Additional models with only a genetic or methylation score were also built. All models were adjusted for age and BMI.

Results:

MS_{SAR} was negatively associated with $ACSA_{VL}$, MVC_{EF} and MVC_{KE}, and explained 10.1%, 35.5% and 40.1% of the variance, respectively. MS_{SAR} explained more variance in these muscular phenotypes than GPS_{SNP}, MS_{SNP} and models including both genetic and methylation scores. MS_{SNP} and GPS_{SNP} accounted for less than 8% and 5% of the variance in all muscular phenotypes, respectively. The genotype and methylation level of *MSTN* was positively related to MVC_{KE} ($p < 0.03$) and explained 12.2% of the variance. The adjusted R^2 and Akaike information criterion showed that models with only a MS_{SAR} performed the best in explaining inter-individual

variance in muscular phenotypes.
Conclusion:
Our results improve the understanding of inter-individual variance in muscular characteristics of older women and suggest a possible application of a sarcopenia-driven methylation score to muscle strength estimation in older women while the combination with a genetic score still needs to be further studied.

Keywords: DNA methylation score, Genetic score, Older women, Model evaluation, Muscle size, Muscle strength

1. Introduction

Muscle mass and strength are two crucial factors in healthy ageing.[81] Older people with lower muscle mass and muscle strength are more likely to have a greater loss of mobility[21] and an increased risk of falls.[22] A ten-year follow-up study by Balogun et al.[273] found that lower-limb muscle mass and muscle strength in older people were positively associated with health-related quality of life.

Many heritability studies have shown a genetic contribution to body composition and muscle strength in older adults. An early twin study on postmenopausal women demonstrated that genetic characteristics account for 52%, 46% and 30% of the variance in lean body mass, leg extensor strength, and grip strength, respectively.[58] An older male twin study conducted by Carmelli et al.[59] showed a decreased genetic association with handgrip strength from 35% to 22% over a 10-year ageing process while the environmental influence increased from 39% to 45%. Furthermore, multiple association studies on athletes, young and old populations have suggested some genetic variants that are closely related to body composition and muscle performance. For example, the D allele of the *ACE* I/D polymorphism is related to higher muscle strength.[274] Older people with the *ACE* DD genotype tend to have greater lean body mass and knee extensor strength than II carriers.[69] The R allele of the *ACTN3* R/X polymorphism is also associated with greater muscle power.[275] Young people with the R allele had significantly higher knee strength and more type IIx fibers than those of XX genotype.[276] The T allele in *FTO* A/T polymorphism is predisposed to increase lean body mass and is more prevalent in elite rugby players, who rely more on appendicular lean mass for success, than other rugby athletes and non-athletes.[71]

To study the combined genetic association with physical phenotypes, a phenotype-driven genetic predisposition score (GPS), which is calculated by adding up the number of predisposing alleles that are positively related to the corresponding phenotype, has been introduced by Williams and Folland.[118] With the application of the phenotype-driven GPS, studies have been able to analyse associations between genetic architectures and physical performance based on multiple polymorphisms. Spanish athletes in endurance activities (e.g. running, road cycling and rowing) were found with a higher endurance-driven GPS than the general population.[119,120] Coronary artery disease patients[123] and older people[212] with higher muscle mass/strength-driven GPS also demonstrated greater muscular improvement after resistance training.

Besides the genetic aspect, muscular phenotypes are also related to multiple external factors such as physical activity and nutrients,[81] which might affect muscle-related gene expression through epigenetic regulation.[277,278] As a link between environment and genes, an epigenetic regulation modifies gene expression through several mechanisms, among which DNA methylation is the one that has been extensively studied. In the human genome, DNA methylation occurs almost exclusively at the 5' position of cytosine in cytosine-phosphate-guanine (CpG) dinucleotides.[88] Many factors such as age, lifestyle and nutrition can trigger DNA methylation changes.[227] DNA methylation in gene promoters is usually associated with a repression of corresponding gene expression,[223] while a recent study by Jeziorska et al.[94] has suggested a positive association between the CpG island methylation in intragenic regions and transcriptional activity. Since DNA methylation is a reflection of environmental exposures and gene expression status, methylation levels of several CpG sites have been suggested as biomarkers for cancer screening[279] and chronological age prediction.[101] A BMI-related epigenetic score developed by Hamilton et al.[128] was found to be associated with body mass, aerobic capacity, type 2 diabetes and cardiovascular disease. The accuracy and sensitivity of diagnostic[126] and prognostic[127] prediction of prostate cancer were also improved with the assistance of DNA methylation scores. Wei et al.[125] built a predictive model for clear cell renal cell carcinoma prognosis based on the methylation of five CpG sites and the model presented reliable predictions across several cohorts. Moreover, DNA methylation scores of specific CpG sites were introduced to the prediction of maternal smoking habit during pregnancy with high accuracy.[132]

In skeletal muscle, epigenetic regulation can be found in development and differentiation processes. The expression of genes from the myogenic regulatory factor and the myocyte enhancer factor families partly rely on DNA methylation to modify skeletal muscle proliferation and differentiation.[280] Meanwhile, some epigenetic traits induced by enviromental stimuli can be maintained for a considerable period (e.g. 30 population doublings of cell culture,[233] seven weeks of detraining[102]), a phenomenon known as “epigenetic memory”.[103] A recent study by Seaborne et al.[102] suggested four genes (*RPL35a*, *UBR5*, *SETD3* and *PLA2G16*) that held epigenetic memory seven weeks after resistance training. All these four genes were characterized by a similar pattern of decreased gene expression with DNA hypermethylation during detraining, and dramatically enhanced gene expression with DNA hypomethylation after retraining.[102] Turner et al.[104] demonstrated five genes (*FLNB*, *MYH9*, *SRGAP1*, *SRGN* and *ZMIZ1*) with increased gene expression in the acute/chronic resistant training and retained hyopmethylation status during seven weeks of detraining, indicating an involvment of these five genes in epigenetic regulation of skeletal muscle characteristics. Lifelong regular physical activity is also associated with hypomethylated promoter regions in genes related to energy metabolism, myogenesis and oxidative stree resistance in ageing muscle.[105] Notably, most methylation studies of skeletal muscle focus on identifying genes with various methylation changes under different intervention phases or between different populations, but the relationship between methylation levels and muscular phenotypes has not been reported.

Several studies have combined genetic and epigenetic profile scores to explore hereditary and environmental associations with physical conditions such as BMI and heart disease risk. Shah et al. found that regression models with only BMI-derived genetic or methylation scores explained less than 10% of the inter-individual variance in BMI, while a model combining both scores improved the explained variance to 13–18%.[130] Another model with integrated genetic and methylation scores also outperformed (13% more accuracy) conventional risk factors in predicting coronary heart disease.[131] Such an approach of combined genetic and epigenetic scores suggests a new approach of studying inter-individual variance and long-term changes in muscle mass and muscle strength. A better understanding of genetic and epigenetic associations with muscular phenotypes can be beneficial to healthy ageing via improved estimation of the probability of muscle degeneration and thus prediction of frailty and sarcopenia. Therefore, our study was conducted to

explore possible genetic and epigenetic connections with muscular phenotypes in a group of older women.

2. Methods

2.1. Participants

Genetic and epigenetic data of 48 older women (aged 65–79 yr) were analysed in this study. These participants were conditionally selected from 247 independently living Caucasian women (aged 65–80 yr) around Manchester Metropolitan University (Crewe, UK), which has been described in details in our previous paper.[281] Briefly, these 48 participants were generally age-matched with no muscular or nervous system problems that would affect their physical performance. With cut-off points of both skeletal muscle index (SMI) less than 6.75 kg/m^2 [16] and hand grip strength (HGS) less than 26 kg (the lower quintile of HGS in all recruited 247 participants), 24 participants were classified as sarcopenic (SMI: 6.00 ± 0.47 kg/m^2, HGS: 23.2 ± 2.5 kg) and the remaining 24 participants were classified as non-sarcopenic (SMI: 7.45 ± 0.67 kg/m^2, HGS: 36.0 ± 3.7 kg). This study followed local ethics approval (Manchester Metropolitan University, Crewe, UK) and consent forms were signed by all participants.

2.2. DNA extraction

A 5 mL venous blood sample was collected from each participant and stored in an EDTA-coated tube at -20 Celsius for DNA extraction. DNA samples were extracted using a QIAcube® and QIAamp® DNA Blood Mini Kit (Qiagen, Crawley, UK) according to the manufacturer's instructions. The extracted DNA samples were stored at -20 Celsius for genotyping and DNA methylation analysis.

2.3. Genotyping

Single nucleotide polymorphisms (SNPs) of seven genes were selected for genotyping. These SNPs have been reported in at least three papers as being related to muscle strength or muscle mass with a consistent direction of favourable alleles (Supplementary Table 1). Duplicate genotyping was firstly made using a 192.24 Dynamic Array® IFC (Fluidigm Corp., South San Francisco, CA, US) and TaqMan SNP genotyping assays (Applied Biosystems, Paisley, UK) following the manufacturer's instructions. Briefly, a genotyping mix (4 μL) consisted of 2 μL assay loading reagent [2x] (Fluidigm), 1 μL SNP genotyping Assay Mix [40X] (Applied Biosystems), 0.2 μL ROX [50X] (Invitrogen, Carlsbad, CA, US) and 0.8 μL DNA-free water (Qiagen). A sample mix (4 μL) contained 1.6 μL DNA samples, 2.0 μL

GTXpress master mix [2X] (Applied Biosystems, PN 4401892), 0.2 µL Fast GT Sample Loading Reagent [20X] (Fluidigm, PN 100–3065), and 0.2 µL DNA-free water. All reaction mixes (7.75 µL, consisting of 3.75 µL genotyping mix and 4 µL sample mix) were loaded onto the Dynamic Array IFC following the manufacturer's instructions. The array was subsequently placed into a thermal cycler (FC1 Fluidigm, PN 100-1279 D1) and the GT 192.24 Fast v1.pcl protocol was performed. The thermal cycling protocol included an amplification at 95 °C for 120 s followed by 45 cycles of denaturation for 2 s at 95 °C and extension for 20 s at 60 °C. Reporter dyes VIC and FAM were used for genotyping based on fluorescence detection.

About 1% of SNP-sample data points showed unsuccessful detection or inconsistent genotype results using the Fluidigm system. These SNP samples were reassessed in duplicates using a StepOnePlus Real-Time PCR system with TaqMan SNP genotyping assays and analysed using StepOnePlus analysis software (Applied Biosystems, version 2.3). The StepOnePlus reaction mix (10 µL) included 0.2 µL DNA sample, 5 µL GTXpress master mix, 4.3 µL nuclease-free water and 0.5 µL TaqMan SNP genotyping assay [20X]. Each reaction mix was amplified for 20 s at 95 °C, followed by 50 cycles of denaturation for 3 s at 95 °C and extension for 20 s at 60 °C. Genotypes were identified based on fluorescence detection of reporter dyes (VIC and FAM).

2.4. DNA methylation analysis

DNA methylation was measured using Illumina® Infinium MethylationEPIC BeadChip arrays (Illumina Inc., San Diego, CA, US) at the Genomics Core facility (Center for Human Genetics, UZ/KU Leuven, Leuven, Belgium). Methylation signal data was read by R 'Minfi' package,[141] background signals were corrected by normal-exponential out-of-band ('Noob') method, and methylation levels (defined as β values, methylation percentages at measured probes) were normalized for blood cell composition by R 'FlowSorted.Blood.EPIC' package. CpG sites were removed from the initial measurement under the following conditions: 1) with a low detection rate ($p > 0.01$ compared with background signal); 2) containing SNPs at the CpG interrogation or at the single nucleotide extension as suggested in the 'Minfi' package (reference array: "Illumina Human Methylation EPIC", annotated by "ilm10b4.hg19"); 3) with cross-reactivity reported in the first supplementary table of Pidsley's study.[143] A final 788,074 CpGs were kept for further analysis.

2.5. Muscular parameters

2.5.1. Biceps brachii thickness

B-mode ultrasonography (7.5 MHz, linear array probe, 38 mm probe length, MyLab®Twice Esaote, Genoa, Italy) was used to measure biceps brachii (BB) thickness (THK_{BB}) on the dominant side (Figure 1a). Participants sat with elbows extended and relaxed. Sagittal plane scans were taken and muscle thickness measured at three sites: 60% of the length from the acromion process of the scapula to the lateral epicondyle of the humerus,[144] and the upper and lower site 1 cm away from the 60%-length site. Muscle thickness was measured using an image processing program (ImageJ, NIH) by the same investigator (intraclass correlation coefficient [ICC] = 0.98, based on duplicate measurements of six participants. The interrater reliability was based on a single scan assessed on two occasions. The following ICC tests were all based on the same participants). The mean muscle thickness of the three sites was recorded as THK_{BB}.

2.5.2. Vastus lateralis anatomical cross-sectional area

With participants in a standing position, the vastus lateralis (VL) origin and insertion were identified at the proximal and distal myotendinous junction under the previously mentioned ultrasound. The VL anatomical cross-sectional area ($ACSA_{VL}$) was measured using an ultrasonography method developed by Reeves et al.[145] with a high reliability and validity compared with magnetic resonance imaging. In brief, participants sat while axial plane scans were taken at 50% muscle length of the VL and recorded in real time, with the ultrasound probe passing over echo-absorptive markers placed over the skin of the VL (as described by Reeves[145]). The acquired images were combined for $ACSA_{VL}$ measurement (Figure 1b). The $ACSA_{VL}$ was measured three times using ImageJ and the mean value was recorded for further analysis. The ultrasound scan was made by the same investigator with good test consistency (ICC = 0.99).

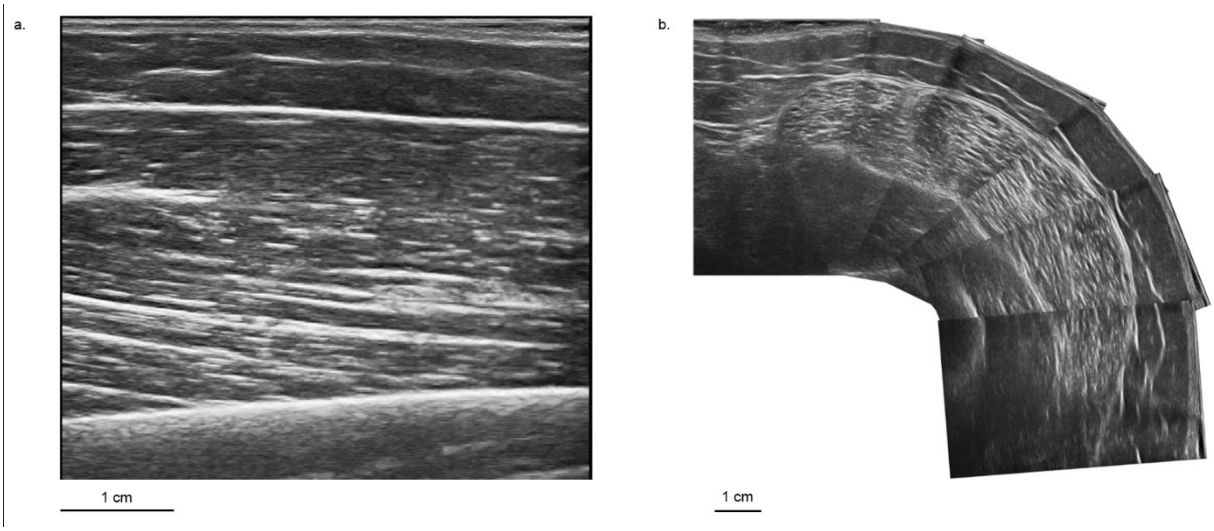


Figure 1. Ultrasound images of the upper arm and the thigh. a. Ultrasound image of the upper arm (Sagittal); b. Ultrasound image of the thigh (cross-sectional).

2.5.3. Maximum isometric elbow flexion torque

Maximum isometric elbow flexion torque (MVC_{EF}) on the dominant side was recorded using a customized dynamometer (MMU, UK), which was calibrated using loads of 0.5-5 kg (with 0.5 kg increments) prior to each strength measurement session. Participants were tested in a seated position with the upper arm parallel to the trunk and the elbow flexed at 60° (0° representing full extension). Participants were asked to hold a force transducer (connected to the dynamometer) and contract their elbow flexors with full effort. Verbal encouragement was given during the test. Three trials were performed with 1 min rest between each trial (ICC = 0.95), with the highest MVC_{EF} used for analysis. Elbow force was recorded at 1,000 Hz and analysed offline at a later date (Labview, National Instruments, Newbury, UK). MVC_{EF} was calculated by the formula: MVC_{EF} = Elbow force × Radius length × cos(30°) with force in N and length in m.

2.5.4. Maximum isometric knee extension torque

Maximum isometric knee extension torque (MVC_{KE}) on the dominant side was recorded using the same system as that used in MVC_{EF} measurement. Participants were tested in a seated position with 60° knee flexion (0° representing full extension). The tested leg was fastened to a force transducer placed 5 cm above the lateral malleolus. Participants were instructed to extend the fastened leg and verbal encouragement was given during the measurement. Three trials were performed with 1 min break between each trial (ICC = 0.96), with the highest MVC_{KE} used for analysis. Knee force was recorded at 1,000 Hz and analysed offline at a later date (Labview, National Instruments, Newbury, UK). MVC_{KE} was calculated by the

formula: MVC_{EF} = Knee force × (Tibia length – 0.05) × cos(30°) with force in N and length in m.

2.6. Statistics, model building and model evaluation

2.6.1. Statistics

SAS 9.4 (SAS Institute, Cary, NC, US) and Python (version 3.7.3) were used for data management and data analysis. Comparisons of muscular phenotypes and methylation scores between the sarcopenic and non-sarcopenic groups were made using independent t-tests. Fisher's exact test was used to compare the distribution of GPS between the two groups. To study combined genetic and epigenetic associations with skeletal muscle, three linear regression models (Models 1–3, Figure 2) were built with muscular phenotypes (THK_{BB}, $ACSA_{VL}$ MVC_{EF} and MVC_{KE}) as dependent variables, and genetic and epigenetic scores as independent variables. Linear models (Models 4–6) with only a genetic or methylation score were also built to study the single genetic or methylation association with muscular phenotypes. All models were adjusted for age and BMI. Data are presented as mean and standard deviation.

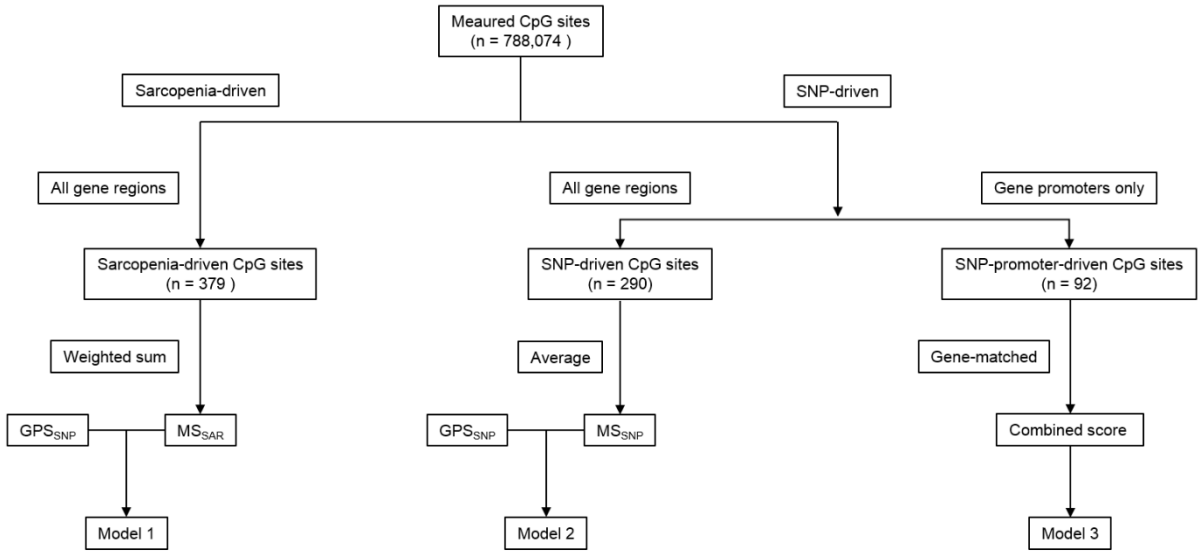


Figure 2. Workflow for model building with combined genetic and methylation scores (Model 1–3). Using muscular phenotypes as dependent variables, Model 1 was built based on methylation levels of CpGs selected from a sarcopenia-driven LASSO regression (MS_{SAR}) and a genetic score calculated by SNP scores of seven muscle-related SNPs (GPS_{SNP}). Model 2 was built from a GPS_{SNP} and a mean methylation score (MS_{SNP}), which was calculated from CpGs located in the genes which contained the seven muscle-related SNPs. Model 3 used a combined score, which

was calculated as the ratio of a SNP score to the promoter methylation level of the corresponding gene, from each of the seven muscle-related SNPs.

2.6.2. Model building

Model 1: muscular phenotypes ~ sarcopenia-driven methylation score (MS_{SAR}) + muscle-driven genetic predisposition score (GPS_{SNP})

This model aimed to analyse the association between muscular phenotypes, muscle-related genetic architecture and sarcopenia-driven methylation levels using a muscle-driven genetic score and a sarcopenia-driven methylation score as independent variables. The least absolute shrinkage and selection operator (LASSO) logistic regression was used for sarcopenia-driven CpG sites selection. The LASSO method combines a linear regression with a L1 penalty on independent variable coefficients to improve prediction accuracy and reduce overfitting.[282] Through a shrinkage parameter tuning, the LASSO method aims to minimize residual sum of squares by setting some coefficients of independent variables to zeros.[282] Therefore, the LASSO method is a powerful tool of selecting strong independent variables from a large set of candidate variables when the amount of independent variables greatly outnumbers the amount of observations.[282] Cross validation is usually used to find an optimal shrinkage parameter.

In the current study, the sarcopenia status was used as the dependent variable (sarcopenia coded as 1 and non-sarcopenia coded as 0) and the methylation levels (β values) at measured CpG sites were used as independent variables. A six-fold cross validation (with the log loss score, the accuracy score and the F1 score as metrics) was used for shrinkage parameter tuning (Figure 3, Supplementary Table 2A). The sarcopenia-driven LASSO regression with an optimal shrinkage parameter selected CpGs (with non-zero coefficients) that were strongly associated with sarcopenia status. The MS_{SAR} was calculated as a weighted sum of the selected CpG methylation levels (the weight for each CpG site was the coefficient from the LASSO regression, Supplementary Table 2B, codes in Supplementary File 1). The "gene symbols" of selected CpG sites were further analysed by gene ontology (GO) and KEGG analysis (databases until June 2019) using Partek Genomics Suite V.7.18 (Partek Inc., St. Louis, MO, US) ("HumanMethylation850" reference, "MethylationEPIC_v-1-0_B4" annotation file, "Homo sapiens" species and hg19 genome build) with a false discovery rate (FDR) control at 0.05.

A summed score of the seven muscle-related SNPs (Supplementary Table 1) was calculated as GPS_{SNP}. Each SNP score was represented by the number of muscle-favourable alleles. For example, the C allele is a muscle-favourable allele in the *ACTN3* rs1815739. Therefore, the SNP score of the *ACTN3* rs1815739 is 2 for a CC genotype, 1 for a CT genotype and 0 point for a TT homozygote.

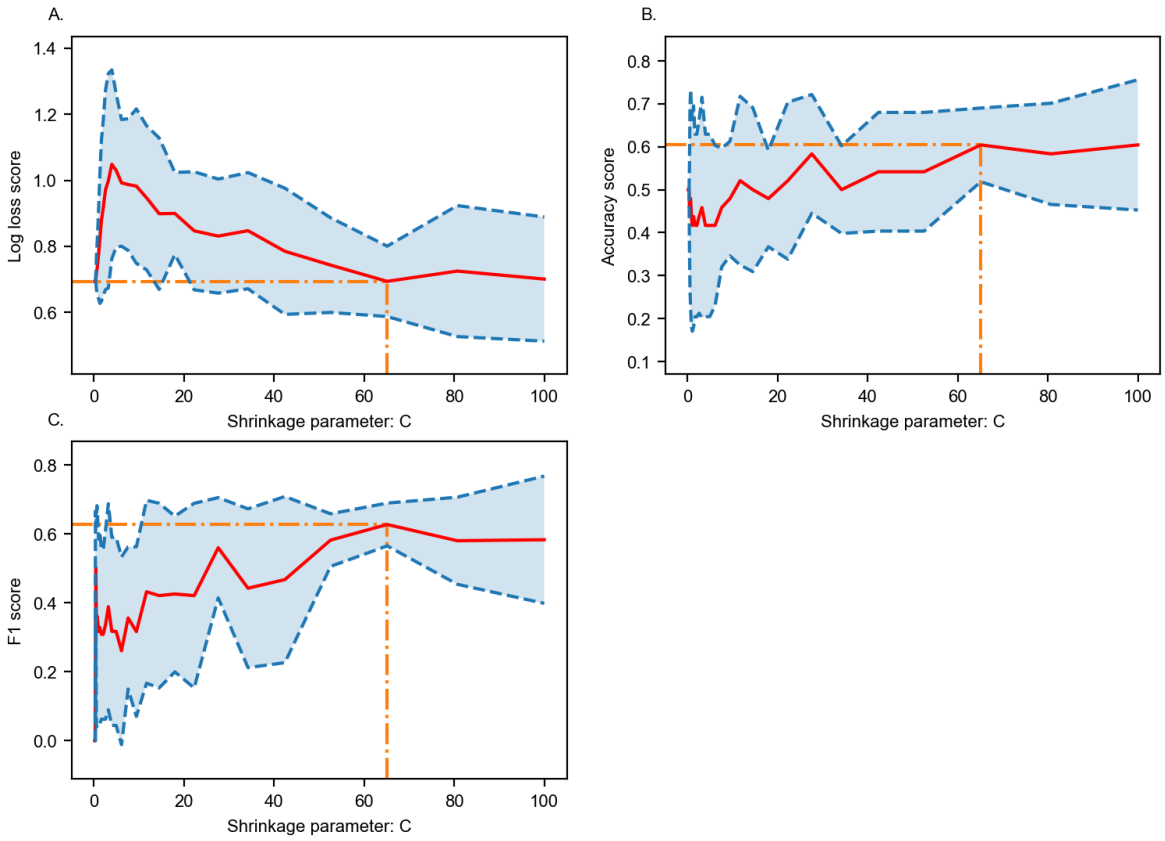


Figure 3. Shrinkage parameter tuning for sarcopenia-driven LASSO logistic regression. The red line represents the mean value of corresponding metric in a six-fold cross validation. The light blue area between the two blue dash lines marks out the range one standard deviation away from the mean value. The orange line demonstrates the optimal shrinkage parameter (C) with the best metric value. a. Changes of the log loss score with the shrinkage parameter. The log loss score reaches an optimal value of 0.69 when C is 65.13; b. Changes of the accuracy score with the shrinkage parameter. The accuracy score reaches an optimal value of 0.6 when C is 65.13; c. Changes of the F1 score with the shrinkage parameter. The F1 score reaches an optimal value of 0.63 when C is 65.13.

Model 2: muscular phenotypes ~ SNP-driven methylation score (MS_{SNP}) + GPS_{SNP}

To evaluate the association between muscle-related genes and muscular phenotypes, this model only included genetic and methylation scores within genes where the seven muscle-related SNPs locate (Supplementary Table 2C). Since

each muscle-related gene contained different amounts of measured CpGs, the mean methylation level of each gene was firstly calculated and the MS_{SNP} was later calculated as the mean of the mean methylation levels of the seven muscle-related genes. Using n^i to represent the number of measured CpGs in the *i*th gene, M_j^i to represent the methylation level of the *j*th measured CpG in the *i*th gene, then the calculation of the MS_{SNP} can be represented as: $MS_{SNP} = (\sum_{i=1}^{7}((\sum_{j=1}^{n^i} M_j^i)/n^i))/7$

Model 3: muscular phenotypes ~ seven gene-wise combined genetic and methylation scores

This model examined each of the seven selected muscle-related gene and studied its association with muscular phenotypes by building a gene-wise combined genetic and methylation profile score. In this model, a methylation score was calculated as the mean methylation level of promoters in each gene because, compared to other gene regions, increased methylation in gene promoters has been more strongly associated with a repression of gene expression.[223] The gene-wise combined score was later calculated as the ratio of a SNP score to the mean methylation level in promoters of the corresponding gene (Supplementary Table 2D) so that a participant with a higher SNP score and a lower methylation score would have a higher gene-wise combined score. For instance, there were five measured CpG sites located in the promoters of *MSTN*. Given that one participant has a *MSTN* SNP (rs1805086) score of 2 and a mean methylation level of 0.32 at the five CpG sites located in *MSTN* promoters, the *MSTN*-wise combined score will be 6.26; if another participant has a *MSTN* SNP score of 1 and a mean promoter methylation level of 0.4, then the *MSTN*-wise combined score will be 2.5. Similar calculations were done in the other six genes and therefore, there were seven gene-wise combined scores (representing each of the seven muscle-related genes) as independent variables in Model 3.

Model 4: muscular phenotypes ~ MS_{SAR}

This model only studied the association between the sarcopenia-driven methylation and muscular phenotypes.

Model 5: muscular phenotypes ~ MS_{SNP}

This model only studied the association between the methylation of muscle-related genes and muscular phenotypes.

Model 6: muscular phenotypes ~ GPS_{SNP}

This model only studied the association between the muscle-related genetic architecture and muscular phenotypes.

2.6.3. Model interpretation and evaluation

Adjusted coefficient of determination (R^2) was used to interpret the explained variance in muscular phenotypes by each linear model. Since a model with more independent variables usually has a higher R^2, the adjusted R^2 is introduced as a modification of the R^2 controlled for the number of independent variables in the corresponding model. In this study, a partial R^2 was also used to illustrate the phenotype variance that an independent variable accounted for in a linear model. The Akaike information criterion (AIC) was used to evaluate the quality of each model with the same muscular phenotype as the dependent variable. The AIC assesses the relative amount of information lost by a given model,[107] therefore, the model with the smallest AIC will be the best model (among all candidate models). Empirically, if another model has an AIC value that is less than two units from the smallest AIC, then that model also has considerable ability to explain variability in the corresponding dependent variable. In that case, more data is needed for model evaluation or a combined model should be created for a better prediction.[283]

3. Results

3.1. Characteristics of participants

Descriptive characteristics of participants are presented in Table 1. Participants in the sarcopenic group had lower body mass (p = 0.003) and BMI (p = 0.005) than the non-sarcopenic group. Values of muscular phenotypes in the sarcopenic group were lower (p < 0.001) than in the non-sarcopenic group, except for THK_{BB} (p = 0.283).

Comparisons of methylation scores are presented in Table 2. The sarcopenic group had a higher MS_{SAR} than the non-sarcopenic group (p < 0.001, Figure 4A) and the combined genetic and methylation score in *VDR* was lower in the sarcopenic group (p = 0.02). The mode and median values of GPS_{SNP} were both 9 in the non-sarcopenic group and the mode and median values of GPS_{SNP} in the sarcopenic group were 7 and 8, respectively. Fisher's exact test for the distribution of GPS_{SNP} between the sarcopenic and non-sarcopenic group showed no difference (p = 0.67, Figure 4C).

Table 1. Characteristics of participants (n = 48)

Participants	Age (year)	Body mass (kg)	Height (m)	BMI (kg/m^2)	HGS (kg)	SMI (kg/m^2)	THK_{BB} (cm)	$ACSA_{VL}$ (cm^2)	MVC_{EF} (N·m)	MVC_{KE} (N·m)
Total (n = 48)	71 ± 4	66.6 ± 12.3	1.59 ± 0.06	26.2 ± 4.5	29.6 ± 7.1**	6.72 ±0.93**	1.82 ± 0.35	16.4 ± 3.9	25.0 ± 5.8**	58.8 ± 19.9*
Sarcopenic (n = 24)	73 ± 4	61.5 ± 9.4**	1.56 ± 0.11	24.4 ± 3.4**	23.2 ± 2.5**	6.00 ± 0.47**	1.76 ± 0.38	14.3 ± 3.0**	21.2 ± 4.8**	44.9 ± 10.2**
Non-sarcopenic (n = 24)	70 ± 3	71.7 ± 12.8	1.60 ± 0.05	28.0 ± 4.9	36.0 ± 3.7	7.45 ± 0.67	1.88 ± 0.33	18.2 ± 3.7	28.7 ± 4.1	71.0 ± 18.3

*: lower than non-sarcopenic group ($p < 0.05$); **: lower than non-sarcopenic group ($p < 0.01$)

Table 2. Description of genetic and methylation profile scores in different groups.

Participants	MS_{SAR}	MS_{SNP}	Gene-wise combined genetic and methylation scores						
			*ACTN3*_combined	*ACE*_combined	*CNTF*_combined	*FTO*_combined	*HIF1A*_combined	*MSTN*_combined	*VDR*_combined
Total (n = 48)	0.02 ± 5.26	0.64 ± 0.01	7.27 ± 5.78	2.73 ± 1.73	2.14 ± 0.46	3.73 ± 2.88	1.02 ± 2.15	2.16 ± 0.02	5.43 ± 2.93
Sarcopenic (n = 24)	5.21 ± 0.53#	0.64 ± 0.01	7.53 ± 6.02	2.68 ± 1.59	2.12 ± 0.49	3.90 ± 3.06	0.91 ± 2.08	2.16 ± 0.02	4.46 ± 2.93*
Non-sarcopenic (n = 24)	-5.17 ± 0.37	0.64 ± 0.01	7.01 ± 5.65	2.78 ± 1.88	2.16 ± 0.44	3.55 ± 2.73	1.13 ± 2.26	2.16 ± 0.02	6.39 ± 2.65

*: lower than non-sarcopenic group ($p < 0.05$); #: greater than non-sarcopenic group ($p < 0.01$)

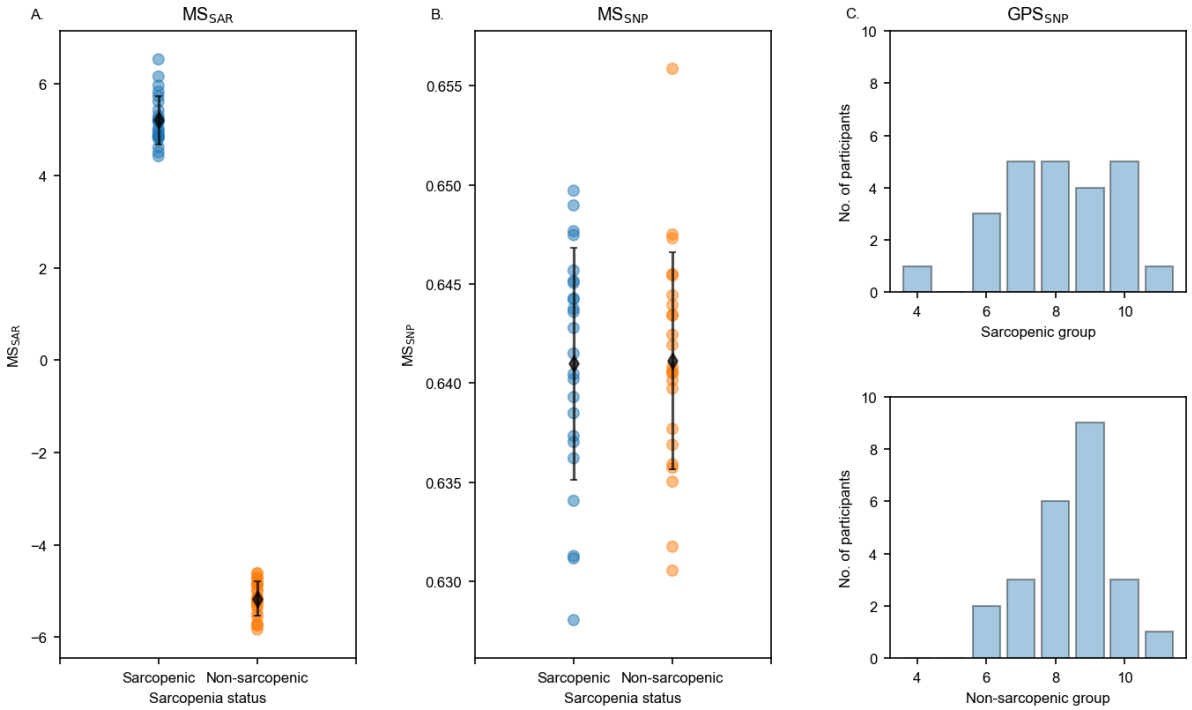


Figure 4. Distribution of genetic and methylation profile scores. (A) MS_{SAR} based on CpGs selected by sarcopenia-driven LASSO regression is higher in the sarcopenic than the non-sarcopenic group ($p < 0.01$). Black diamond markers represent the mean MS_{SAR} in each group. (B) MS_{SNP} is not different between the sarcopenic and non-sarcopenic groups ($p = 0.96$). (C) Fisher's exact test for the distribution of GPS_{SNP} showed no difference between the sarcopenic and non-sarcopenic group ($p = 0.67$).

3.2. CpG sites selected from the sarcopenia-driven LASSO logistic regression

379 CpG sites were selected from the sarcopenia-driven LASSO logistic regression, indicating a possible association between these CpGs and sarcopenia. These selected CpGs located in 190 genes, with the *PIWIL1* gene contributing most (n = 4) CpGs (Supplementary Table 3A). GO analysis on the identified genes showed that 29 GO terms were enriched after FDR control (q value < 0.05, Supplementary Table 3B). Many of these GO terms were associated with protein binding (e.g. antigen binding and cell adhesion), MHC protein complex, signal transduction (e.g. receptor binding and transport vesicle membrane) and synapse structure (e.g. synapse assembly and synapse organization). The most enriched GO term was the "peptide antigen binding" term (q value = 0.002), which included four hypermethylated (higher methylation in the sarcopenic group than non-sarcopenic group) CpGs and three hypomethylated (lower methylation in the sarcopenic group) CpGs (Supplementary Table 3C). KEGG analysis showed 46 pathways that were

enriched (q value < 0.05, Supplementary Table 3D). Many pathways were related to immune system function (e.g. allograft rejection and T helper cells differentiation) and diseases (e.g. autoimmune thyroid disease and viral myocarditis), and chronic disorders (e.g. type I diabetes mellitus and rheumatoid arthritis). The most enriched pathway was the “Asthma” pathway with three hypermethylated CpGs and four hypomethylated CpGs (Supplementary Table 3E).

3.3. Muscular phenotypes with genetic and methylation scores

Main results of linear regression models with both genetic and methylation scores are presented in Table 3 (complete results in Supplementary Table 4). In Model 1, the MS_{SAR} was negatively related to MVC_{EF} and MVC_{KE} ($p < 0.01$) and explained 33.2% and 39.4% of the variance, respectively. With one unit increase in the MS_{SAR}, MVC_{EF} and MVC_{KE} decreased by 0.67 and 2.63 N·m, respectively. The GPS_{SNP} was not significantly associated with any muscular phenotypes. In Model 2, neither the MS_{SNP} nor the GPS_{SNP} was significantly correlated to muscular phenotypes. In Model 3, only the combined genetic and methylation score in the *CNTF* was positively related to MVC_{KE} ($p = 0.03$) and explained 12.2% of the MVC_{KE} variance. A one-score addition in the *CNTF* combined score was associated with 15.7 N·m increase in MVC_{KE}.

Results of linear models with only a genetic/methylation score are presented in Table 4 (complete results in Supplementary Table 4). The MS_{SAR} alone (Model 4) was negatively associated with $ACSA_{VL}$, MVC_{EF} and MVC_{KE}, and explained 10.1%, 35.5% and 40.1% of the variance, respectively. The MS_{SNP} and GPS_{SNP} were not associated with any muscular phenotypes. Specifically, the MS_{SNP} explained less than 8% of the variance in muscle size and less than 1% of the variance in muscle strength. The GPS_{SNP} accounted for less than 5% of the variance in all muscular phenotypes.

Explained variance of muscular phenotypes by the six models are presented in Table 5. Model 1 with both the MS_{SAR} and GPS_{SNP} explained less phenotype variance than Model 4, which included only an MS_{SAR}, and more variance in muscle strength than Model 6, which included only a GPS_{SNP}. Model 2 with the MS_{SNP} and GPS_{SNP} explained less variance in muscle size than Model 5. When compared with Model 6, Model 2 explained more variance in muscle size but less variance in muscle strength. Models with an MS_{SAR} (Model 1 and 4) explained more variance in muscle strength (MVC_{EF} and MVC_{KE}) than models without MS_{SAR}. When comparing

models with genetic and methylation profile scores within the pre-selected seven muscle-related genes, Model 3 explained less variance than Model 2 in all muscular phenotypes except for MVC_{KE}. Notably, the explained variance in muscle size (THK_{BB} and $ACSA_{VL}$) was similar across all models. This was possibly because BMI was closely related to muscle size and explained a considerable percentage (20.1% to 61.4%) of the variance (Table 3 and 4).

In the aspect of model evaluation, Model 5, which included only an MS_{SNP}, explained the most variance in THK_{BB} (with the highest adjusted R^2 value) and outperformed other models in the prediction of THK_{BB} (with the lowest AIC value). Model 4, which includes only an MS_{SAR}, explained the most variance in $ACSA_{VL}$, MVC_{EF} and MVC_{KE} (with the highest adjusted R^2 values), and performed better than other models in the prediction of those muscular phenotypes ($ACSA_{VL}$, MVC_{EF} and MVC_{KE}) (with the lowest AIC values). Notably, the AIC differences were smaller than two between Model 2 and 5 (in THK_{BB}), and Model 1 and 4 (in $ACSA_{VL}$, MVC_{EF} and MVC_{KE}), indicating that more data might be needed before deciding if GPS_{SNP} should be included into a model.

Table 3. Main results of linear models (Model 1-3) with combined genetic and methylation scores

		Model 1				Model 2				Model 3		
		MS_{SAR}	GPS_{SNP}	Age	BMI	MS_{SNP}	GPS_{SNP}	Age	BMI	*CNTF*_combine	Age	BMI
THK_{BB}												
	Coef	<0.01	-0.01	0.03	0.04	-16.47	-0.02	0.04	0.04	0.09	0.03	0.04
	Partial R^2	0.002	0.003	0.145	0.201	0.080	0.007	0.198	0.270	0.017	0.142	0.258
	p	0.76	0.74	**0.01**	**<0.01**	0.07	0.60	**<0.01**	**<0.01**	0.44	**0.02**	**<0.01**
$ACSA_{VL}$												
	Coef	-0.16	0.16	-0.02	0.57	-97.93	0.26	-0.02	0.64	1.33	-0.04	0.73
	Partial R^2	0.080	0.008	0.001	0.499	0.041	0.023	0.001	0.591	0.052	0.004	0.614
	p	0.07	0.56	0.89	**<0.01**	0.20	0.34	0.84	**<0.01**	0.18	0.72	**<0.01**
MVC_{EF}												
	Coef	-0.67	0.29	-0.05	0.08	-23.51	0.78	-0.27	0.37	2.02	-0.24	0.42
	Partial R^2	0.332	0.009	0.001	0.006	0.001	0.042	0.033	0.086	0.023	0.026	0.094
	p	**<0.01**	0.55	0.81	0.62	0.88	0.18	0.23	**0.05**	0.36	0.33	0.06
MVC_{KE}												
	Coef	-2.63	-0.37	-0.58	-0.53	193.22	1.56	-1.30	0.72	15.68	-0.99	1.70
	Partial R^2	0.394	0.001	0.019	0.021	0.003	0.014	0.053	0.028	0.122	0.042	0.142
	p	**<0.01**	0.82	0.38	0.36	0.74	0.45	0.14	0.29	**0.03**	0.22	**0.02**

Table 4. Main results of linear models (Model 4-6) with only genetic or methylation scores

		Model 4			Model 5			Model 6		
		MS_{SAR}	Age	BMI	MS_{SNP}	Age	BMI	GPS_{SNP}	Age	BMI
THK_{BB}										
	Coef	<0.01	0.03	0.04	-15.80	0.04	0.04	-0.01	0.03	0.04
	Partial R^2	0.001	0.147	0.211	0.075	0.194	0.272	0.002	0.147	0.248
	p	0.82	**0.01**	**<0.01**	0.07	**<0.01**	**<0.01**	0.80	**0.01**	**<0.01**
$ACSA_{VL}$										
	Coef	-0.17	<0.01	0.57	-107.42	<0.01	0.64	0.31	-0.07	0.64
	Partial R^2	0.101	<0.001	0.496	0.048	<0.001	0.583	0.031	0.009	0.579
	p	**0.04**	1.00	**<0.01**	0.16	0.98	**<0.01**	0.26	0.54	**<0.01**
MVC_{EF}										
	Coef	-0.69	-0.02	0.06	-51.35	-0.22	0.34	0.79	-0.28	0.37
	Partial R^2	0.355	<0.001	0.004	0.003	0.022	0.074	0.044	0.036	0.087
	p	**<0.01**	0.90	0.69	0.73	0.33	0.07	0.17	0.21	**0.05**
MVC_{KE}										
	Coef	-2.60	-0.61	-0.51	115.81	-1.17	0.68	1.45	-1.21	0.72
	Partial R^2	0.401	0.022	0.020	0.001	0.044	0.026	0.013	0.051	0.029
	p	**<0.01**	0.34	0.37	0.84	0.18	0.31	0.47	0.15	0.28

Table 5. Adjusted R^2 and AIC of linear models

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
THK_{BB}							
	Adj R^2	0.240	0.299	0.255	0.256	0.311	0.256
	AIC	-103.5	-107.2	-100.4	-105.4	-108.9	-105.4
$ACSA_{VL}$							
	Adj R^2	0.585	0.568	0.561	0.592	0.568	0.560
	AIC	86.4	88.3	92.9	84.8	87.3	88.1
MVC_{EF}							
	Adj R^2	0.371	0.060	-0.013	0.380	0.042	0.081
	AIC	148.4	167.3	174.8	146.8	167.3	165.3
MVC_{KE}							
	Adj R^2	0.392	<.001	0.112	0.406	0.010	0.021
	AIC	251.4	273.8	272.5	249.5	272.5	272.0

3.4 Correlation and regression analysis on actual and predicted values inferred by Model 4

As Model 4 is the most powerful model in predicting muscle size and strength than the rest models in Table 5, correlation and linear analysis were made between actual and predicted values inferred by Model 4 (Figure 5, Supplementary Table 4F). Moderate associations were found between the actual and predicted values with correlation coefficients ranging from 0.55 to 0.79. Notably, there are two distinct clusters in the plots of MVC_{EF} and MVC_{KE}. This is because the MS_{SAR} was strongly related to the two muscular phenotypes and accounted for a considerable percent of the variance in the corresponding phenotype (Table 4). Since the sarcopenic participants received positive MS_{SAR} and the non-sarcopenic participants had negative MS_{SAR}, the average MVC_{EF} and MVC_{KE} differences between the sarcopenic and non-sarcopenic group led by the MS_{SAR} difference were 7 N·m and 27 N·m, respectively, while negligible average differences (0.1 N·m in MVC_{EF} and 0.9 N·m in MVC_{KE}) were attributable to age and BMI. Therefore, the two clusters in the plots of MVC_{EF} and MVC_{KE} actually represent the sarcopenic and non-sarcopenic groups.

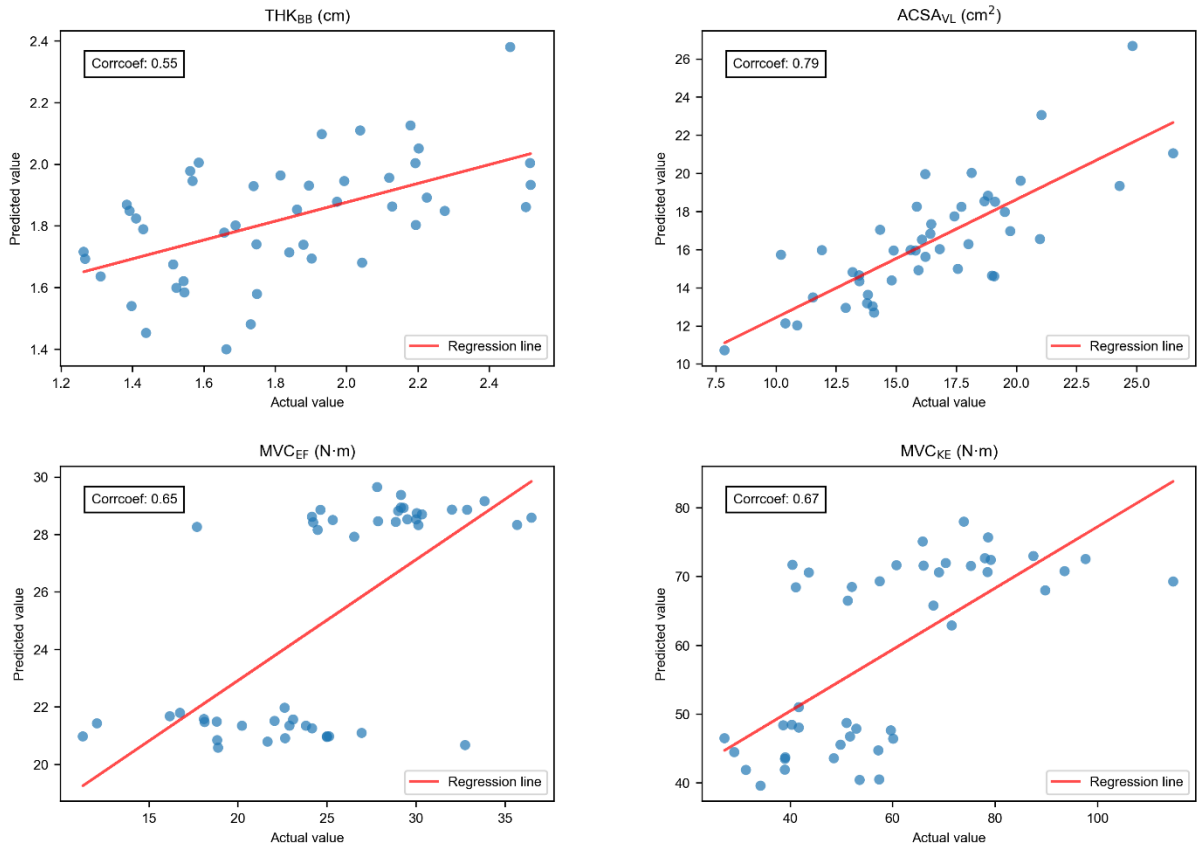


Figure 5. Plots of actual and predicted values of muscular phenotypes inferred by Model 4. Actual and predicted values are moderately correlated with coefficients

ranging from 0.55 to 0.79. The two distinct clusters in the plots of MVC_{EF} and MVC_{KE} are due to the difference of MS_{SAR} in the sarcopenic and non-sarcopenic groups.

4. Discussion

The current study explored the association between muscular phenotypes, genetic architecture and DNA methylation via linear regression models in sarcopenic and non-sarcopenic elderly women. Genetic architecture was represented as a GPS that was calculated from seven muscle-related SNPs. The DNA methylation was represented as either a sarcopenia-driven methylation score, which was calculated as a weighted sum of the methylation levels of 379 sarcopenia-driven CpG sites, or a gene-wise methylation score, which was calculated as an average of the methylation levels within muscle-related genes. Based on the six linear models used in this study, the sarcopenia-driven methylation score was negatively related to $ACSA_{VL}$, MVC_{EF} and MVC_{KE}, and explained more variance in these muscular phenotypes than the GPS, the gene-wise methylation score and the models with combined genetic and methylation scores. The adjusted R^2 and AIC showed that models with only a methylation score had the best performance in explaining inter-individual variance in muscular phenotypes while more data are needed to determine the inclusion of GPS into the models. Moreover, the model with gene-wise combined genetic and methylation scores demonstrated that the genotype and methylation level in *CNTF* was closely related to knee extensor strength, indicating a close association between *CNTF* and knee strength.

4.1. Sarcopenia-driven CpG sites

DNA methylation changes have been examined in studies of ageing and resistance training, but no study has reported the DNA methylation association with sarcopenia except for our own work.[281] Using the sarcopenia-driven LASSO logistic regression, our study identified 379 CpG sites that were possibly related to sarcopenia. Zykovich et al.[101] identified 5,963 CpG sites that were related to ageing based on skeletal muscle tissue. Bell et al.[230] found 490 ageing-associated CpGs from blood samples. However, none of those ageing-related CpGs were found among the sarcopenia-driven CpGs identified in our study. Seaborne et al. studied DNA methylation changes in skeletal muscle during resistance training and identified 2,445 CpG sites that were differentially methylated after a seven-week loading stimuli and 1,883 CpGs that were association with an unloading phase. We shared one CpG site in each of the loading and unloading phase, with both CpGs located in the intergenic

region (Supplementary Table 5A). We further compared our CpGs with those identified by Turner et al.,[104] who analysed transcriptome and methylome associations after acute/chronic resistance training, but no common CpG was found. Notably, Turner et al. reported three genes (*ETF1*, *ETV1* and *SH3KBP1*) that were up-regulated after acute/chronic resistance training[104] while some hypermethylated sarcopenia-driven CpGs identified in our study were found to locate in promoter regions of those three genes (Supplementary Table 5B).

The gene *ETF1* is a member of the human transcriptional enhancer family. Recent research on human liver HepG2 cell line showed that the *ETF1* gene was involved in the regulation of transcript stability.[284] The gene *ETV1* is involved in multiple cellular activities that are related to physical performance. *ETV1* knockout mice demonstrated abnormal cardiac conduction[285] and neuromuscular impairment.[286] The gene *SH3KBP1* belongs to a gene group of putative motility modifiers, and the knocking down of *SH3KBP1* leads to reduced cell migration in scratch wound assays.[287] Since it has been established that hypermethylated gene promoters are associated with repressed gene expression,[288] the identification of hypermethylated CpGs in promoters of these three genes (*ETF1*, *ETV1* and *SH3KBP1*) in our study indicates possible down-regulated cellular activity in association with sarcopenia.

4.2. Evaluation of linear models

In our study, the sarcopenia-driven methylation score (MS_{SAR}) was closely related to muscle strength and explained 33.2% to 40.1% of inter-individual variance in all models (Model 1 and 4). This indicates a possible application of the MS_{SAR} to the estimation of skeletal muscle strength in older women. Meanwhile, we should be aware that the participants in this study belong to two groups (i.e. sarcopenic and non-sarcopenic groups) which have significant difference in muscle strength (Table 1) and MS_{SAR} (Table 2). Therefore, when applying the MS_{SAR} to a population with less variability in muscle strength (e.g. a group with only physically fit older people or a group with only sarcopenic participants), the corresponding muscular variance explained by the MS_{SAR} might decrease to some extent. In fact, the MS_{SAR} was found to explain less than 8% of the muscular variance within the sarcopenic or non-sarcopenic group when analysed separately (Supplementary Table 4E). Clearly, future studies on larger cohorts are still needed to evaluate the feasibility of applying the MS_{SAR} for muscle strength evaluation.

Additionally, we found that genetic profile scores based on seven selected genes (GPS_{SNP}) explained up to 4.4% of the variance in muscle size and strength, methylation levels in the seven selected genes (MS_{SNP}) explained up to 8% in the studied phenotypes, while the MS_{SAR} explained 10.1–40.1% of the individual differences in muscle size and strength in our sample of older women. These results showed that genetic and methylation profiles on several representative genes only explained limited muscular variability. Moreover, by comparing the AIC, the model with only an MS_{SAR} showed the best performance in explaining the variance in muscle size and strength. This, again, indicated that using the data from a small set of representative genes might not well explain muscular variability. Admittedly, the GPS_{SNP} in this study is based on only seven SNPs, however, the individual muscular variance explained by genetic structures might still have limited improvement even with an increased number of candidate SNPs. Previous studies have demonstrated that even based on a larger candidate pool of more than one hundred genetic variants, the data-driven GPS only explained up to 7% of the variance in muscle mass and strength.[124,212] Therefore, it is possible that the genetic architecture only accounts for a small portion of muscular variability during ageing – or we have not yet used the optimal methodology to include all contributing genetic factors, while a larger proportion of the variance is taken up by DNA methylation. Since DNA methylation is representing the sum of short-term and long-term environmental factors, the finding that methylation levels explains a larger proportion of the variance in muscle morphology and strength than genetic profiles might indicate that environmental elements account for more variance than genetic factors in skeletal muscle during ageing, which is in line with the findings of previous heritability studies.[59,61,289]

Notably, the model with gene-wise combined genetic and methylation score (Model 3) showed that the genotype and methylation level in *CNTF* was closely related to knee extensor strength. The *CNTF* gene encodes ciliary neurotrophic factor, a polypeptide that promotes neuronal cell differentiation and neurite outgrowth, and exerts a neuroprotective effect by preventing motor neuron degeneration.[290] Indeed, *CNTF* G allele carriers have shown higher knee strength than A allele homozygotes at both slow and fast contraction speed across a large age span (20–90 yr).[291] Our finding provides supportive evidence for the association between *CNTF* and knee strength.

4.3. Limitations

We acknowledge that, despite presenting signficant associations, our study has a limited sample size. Because of the limited sample size, we could only use adjusted R^2 and AIC for model evaluation. Therefore, our results still need to be examined in different cohorts with large sample sizes. Moreover, DNA methylation is only one mechanism of epigenetic regulation. Future models including other epigenetic mechanisms (e.g. histone methylation and acetylation), genome conformation, and transcriptome analysis might make the model more reliable.

Another limitation is that the methylation data used in this study was based on blood samples. It is well known that DNA methylation is tissue-specific,[95] so the methylation data obtained from blood might not fully represent the methylation status in other tissues. Although venous blood is more easily obtained, methylation status could be more informative if DNA was derived from skeletal muscle tissue. Meanwhile, the Illumina MethylationEPIC BeadChip used for methylation analysis in our study only covers 850K CpG sites, which is a small proportion of the 28 million CpG sites in the human genome, and is not always informative since many CpG sites are omitted.[143] Therefore, a methylome-wide association study should be more powerful in identifying sarcopenia-driven or muscle-related CpG sites for model building.

5. Conclusions

Our study combined genotypes and DNA methylation levels to evaluate their associations with muscle size and strength in older women. We found that a sarcopenia-driven methylation score explained more inter-individual variance in muscle strength and thigh muscle size than a genetic score or models with both genetic and methylation scores. Our results suggest a possible application of a sarcopenia-driven methylation score to identify older adults who are at risk of muscle weakness conditions (e.g. sarcopenia and frailty) using routine blood samples, while the combination with a genetic score still needs to be further studied.

6. Conflict of interest

The authors declare no conflicts of interest. The authors certify that they comply with the ethical guidelines for publishing in the Journal of Cachexia, Sarcopenia and Muscle: update 2017.[272]

7. Acknowledgements

This study is supported by the Research Foundation – Flanders (FWO) [grant G.0898.15], and a PhD grant to Lingxiao He and Praval Khanal by the Erasmus Mundus ‘Move-AGE’ joint doctorate program.

PART 3

SUMMARY AND GENERAL DISCUSSION

1. Summary

Inter-individual differences in muscle mass and strength are observed during the ageing process, when muscles adapt to exercise training and after the cessation of exercise. Muscle mass and strength phenotypes are found to be partly related to DNA sequence variation and epigenetic regulation, with DNA methylation as the most studied mechanism. However, few studies have been done to unravel muscle-related inter-individual differences among older adults with focus on genetic variation and DNA methylation. Insights in the genetic associations and methylation patterns with inter-individual differences in ageing muscle will enhance our understanding of muscle degeneration during the ageing process and can be helpful for the identification of individuals at risk of muscle weakness conditions (e.g. sarcopenia and frailty).

Therefore, the objective of this thesis was to explore the exploratory power of sets of genetic sequence polymorphisms and methylation information in ageing muscle. Those genetic and methylation markers are firstly identified based on a list of candidate gene polymorphisms and genome-wide methylation data (**paper 1**, **2**, and **4**), and are subsequently transformed into a genetic/methylation profile score to explore their underlying relationships with ageing muscle. To complete these goals, this thesis firstly studied the genetic association with baseline muscle mass and strength in older people (**paper 1**), and their responses in muscular phenotypes after one year of supervised training (**paper 1**). Gaining new insights into the contribution of genetic sequence variation in the variability in detraining responses one-year after the cessation of training was the objective of **paper 2**. Based on blood methylation profiles, the thesis further identified DNA methylation differences that were associated with muscular variance among older women (**paper 3**). Finally, genetic and methylation profiles were combined to study their associations with muscle morphology and strength in older women (**paper 4**). D-Table 1 summarises the hypotheses and results of each paper.

D-Table 1. Summary of hypotheses/research questions and results of each paper

Paper	Hypotheses/Research questions	Results
1	• Data-driven GPS has a significantly predictive value in baseline muscle mass and strength in older adults. An individual with a higher GPS will have a larger muscle mass and strength. • Muscle mass and strength will increase after training, however, substantial inter-individual variability in these muscular response phenotypes exists. • Data-driven GPS is positively associated with muscle adaptations to exercise. An individual with a higher GPS will have a larger muscular improvement after structured training.	• Data-driven GPS was positively related to isometric knee extensor strength at baseline and explains 3.2% of the variance. A one-unit increase in GPS leaded to 4.73 Nm increase in knee strength. (Hypothesis **confirmed**) • Muscle mass and knee strength increased significantly after one-year of training with obvious inter-individual variance of -16–22% in muscle mass and -21–59% in knee strength. (Hypothesis **confirmed**) • Data-driven GPS was positively related to muscle mass and knee extensor strength changes induced by exercise. GPS explained 14% and 27% of the variance in muscle mass and knee strength, respectively. In response to the training program, with one-unit increase in GPS, muscle mass and knee strength increased by 1.78% and 3.86%, respectively. (Hypothesis **confirmed**) • Four out of the 19 identified SNPs were significantly associated with gene expression in skeletal muscle based on the GTEx database[292] and three out of the four SNPs showed the same expression direction as that

		reported in the GTEx database. These three SNPs were involved in domains of muscle growth and muscle structure. (**Additional** findings)
2	• Muscle mass and strength will decrease after a one-year cessation of training with inter-individual differences. • Data-driven GPS is negatively related to the decline in muscle mass and strength after the cessation of training. An individual with a higher GPS will have smaller loss in muscle mass and strength.	• Isotonic and isokinetic knee extensor strength decreased, but muscle mass and isometric knee extensor strength did not decrease after one-year cessation of training. (Hypothesis **partially confirmed**) • Inter-individual variance was found in changes of muscle mass and knee strength. The change in muscle mass ranged from -14% to 35%. The change in isometric knee strength was -32% to 41%. The peak speed of isotonic knee movement showed a change range of -22–13%, and the isokinetic knee strength had a change range between -29% and 21%. (Hypothesis **confirmed**) • A data-driven GPS was closely related to changes in muscle mass and strength after the cessation of training, and explained 26–37% of the variance. Participants with a higher GPS had smaller losses in muscular phenotypes. (Hypothesis **confirmed**) • 23 out of the 46 identified SNPs were closely associated with gene expression in skeletal muscle based on the GTEx database and 8 out of the 23 SNPs showed the same expression direction as that reported in the GTEx database. The 8 SNPs were involved in domains of muscle

		growth, metabolism, DNA methylation and neural control. (**Additional** findings)
3	• To identify sarcopenia-related genes and CpGs that demonstrate differentially methylated patterns between sarcopenic and non-sarcopenic older women. • To explore possible sarcopenia-related biological functions and pathways based on sarcopenia-related genes.	• 6,258 CpGs were differentially methylated ($p < 0.01$) between generally age-matched sarcopenic and non-sarcopenic women. The sarcopenic group had higher methylation levels in gene promoters and lower methylation levels in gene bodies (Hypothesis **confirmed**) • Differentially methylated genes are involved in multiple pathways that are related to muscle function, actin cytoskeleton regulation, energy metabolism and signal transduction. (Hypothesis **confirmed**)
4	• Both genetic and methylation profile scores have a significant predictive value to explain muscle size and strength variability in older women. • A model with both genetic and methylation profile scores explains more inter-individual variability in muscle morphology and strength than a model with only a genetic or methylation profile score.	• GPS_{SNP} was not related to muscular phenotypes. MS_{SAR} was related to vastus lateralis size, elbow and knee strength. (Hypothesis **partially confirmed**) • A model with only an MS_{SAR} explained the highest proportion of variability in muscle morphology and strength in older women compared to other models: either a model with only a GPS_{SNP} or an MS_{SNP}, or models containing both genetic and methylation profile scores. (Hypothesis **rejected**)

2. General discussion

Regardless whether this thesis was based on two projects, which had different participants and candidate gene sets, we will explore the associations of genetic and methylation markers with individual variability in ageing muscle characteristics based on the results of each paper and the comparisons between them.

2.1. Genetic association with *baseline* muscle mass and strength in older people

Heritability studies have shown that a considerable proportion of muscle mass and strength can be explained by genetic factors. Abney et al.[293] conducted a study in a founder population with extensive genealogical records, which increased the power to detect dominance genetic variance, and found that genetic factors contributed up to 76% of the variance in fat free mass. Arden et al.[58] studied 353 pairs of postmenopausal twins and reported that the explained variance by genotypes was 52% in lean body mass, 30% in grip strength, and 46% in leg extensor power. Carmelli et al.[59] conducted a 10-year follow-up study in 152 pairs of older male twins and reported that genes accounted for 35% of the grip strength variance at baseline and 22% of the variance 10 years later. They further analysed the genetic association with hand grip strength change during the 10-year period and demonstrated a genetic contribution of 35% to the variance in strength loss. Frederiksen et al.[60] studied 1,757 twin pairs aged 45–96 years and found a 52% heritability in grip strength. Besides heritability studies on grip strength, the genetic effect on low limbs was also reported. In a study of Finnish twin older women (aged 63–96 yrs), Tiainen et al.[289] found 31% of the knee extensor strength was explained by genes. The same research team also studied female twins (aged 63–76 yrs) over a 3-year follow-up and found that the genetic effect on muscle strength decreased with age. Genetic factors were found to explain 58% of the variance in knee extensor strength at baseline and 56% at follow-up level, and 67% of the variance in knee extensor power and 48% of the variance after three years.[61]

Compared to reported heritability estimates, studies testing a genetic predisposition score (GPS) often demonstrate only limited power to explain variability in muscular phenotypes. Charlier et al.[124] analysed muscular phenotypes of 565 adults (aged 19–73 yrs) by GPS models built on 153 muscle-related single nucleotide polymorphisms (SNPs). The explained variance by GPS was 2.9–6.1% in muscle mass, 3.8–5.4% in isometric knee extensor strength, 2.2–6.8% in isotonic knee strength and 3.5–6.8% in isokinetic knee strength.[124] Similarly, **paper 1** in this thesis

also showed that only 0.7% of the variance in muscle mass and 3.2% of the variance in isometric knee extensor strength could be explained by data-driven GPSs (based on 170 candidate SNPs). The total GPS calculated from seven muscle-related SNPs (*ACTN3* rs1815739, *ACE* rs4341, *CNTF* rs1800169, *FTO* rs9939609, *HIF1A* rs11549465, *MSTN* rs1805086 and *VDR* rs2228570) contributed to 0.2–3.1% of the variance in muscle size and 1.3–4.4% of the variance in muscle strength in older women using a genetic model (Model 6) of **paper 4**. Using the shared five SNPs (*ACTN3* rs1815739, *CNTF* rs1800169, *MSTN* rs1805086, *FTO* rs9939609 and *HIF1A* rs11549465) in **paper 1** and **4**, GPSs were calculated with the predisposing alleles as defined in **paper 4** (P4-Supplementary Table 1). As presented in D-Table 2, those GPSs only explain up to 3.7% of the variance in both muscle mass and knee extensor strength at baseline.

D-Table 2. Genetic association with muscle mass and knee strength (based on five shared SNPs in **paper 1** and **4**)

Paper	Paper 1								Paper 4							
	SMM_baseline (kg)				PT_{IM60}_baseline (Nm)				SMM (kg)				PT_{IM60} (Nm)			
Parameter	Estimate	β value	Partial r^2	*p*	Estimate	β value	Partial r^2	*p*	Estimate	β value	Partial r^2	*p*	Estimate	β value	Partial r^2	*p*
GPS	-0.45	-0.09	0.037	0.007	-4.53	-0.14	0.037	0.009	0.22	0.07	0.007	0.586	0.18	0.01	<0.001	0.955
SEX (M=1, F=0)	11.66	0.90	0.802	<0.001	56.95	0.69	0.497	<0.001	-	-	-	-	-	-	-	-
AGE	-0.10	-0.07	0.024	0.030	-2.45	-0.28	0.144	<0.001	-0.10	-0.13	0.024	0.306	-1.12	-0.21	0.044	0.178
BMI	0.15	0.08	0.034	0.010	0.05	<0.01	<0.001	0.934	0.35	0.53	0.282	<0.001	0.69	0.16	0.026	0.300
Intercept	22.89				300.97				13.89				119.54			
Adj. r^2	0.803				0.525				0.257				0.009			

D-Table 3. Comparisons of significant SNPs contributing to data-driven GPSs at different timepoints

Timepoints	SNP	Gene	SMM baseline	PT_{IM60} baseline	ΔSMM training	ΔPT_{IM60} training	ΔSMM detraining	ΔPT_{IM60} detraining	ΔPT_{IK60} detraining	ΔPT_{IK240} detraining
Baseline and Training										
	rs3110697	*IGFBP3*	X	X		X				
Baseline and Detraining										
	rs10497520	*TTN*		X					X	
	rs3797297	*FST*		X						X
Training and Detraining										
	rs2390760	*METTL21C*			X		X	X	X	
	rs3762546	*MSTN*			X		X	X		X
	rs7703033	*MTRR*				X			X	

The gap in explained variance between heritability and measured genetic variant studies, known as the problem of *missing heritability*, has been fuel for discussions for more than a decade.[294] One possible explanation is the incompleteness of the candidate genetic variants set used in a GPS study. For instance, the GPS studies in the current thesis did not include all muscle-related genetic variants. In recent years, several genome-wide association studies (GWASs) have identified multiple SNPs that are closely related to muscle strength and mass. Matteini et al.[295] conducted a meta-analysis of GWASs on handgrip strength among older Europeans aged over 65 yrs and identified two significant (p-value < 5×10^{-8}) and 39 suggestive (p-value < 5×10^{-5}) SNPs that were associated with grip strength. Willems et al.[191] identified 16 loci associated with grip strength (p-value < 5×10^{-8}) based on a large-scale genetic analysis and those loci were involved in multiple pathways such as myofiber function, neuronal maintenance and signal transduction. Tikkanen et al.[110] conducted a GWAS of grip strength and identified 101 loci (p-value < 5×10^{-8}) which explained 1.5% of the variance in grip strength. The same research team further completed a meta-analysis of the discovery GWASs and identified 139 grip strength-associated loci, which explained 1.7% of the variance in grip strength.[110] In a GWAS on one thousand Americans, Liu et al.[111] identified two genome-level significant SNPs (rs16892496 and rs7832552), and another 146 suggestive (p-values < 1.26×10^{-4}) SNPs associated with lean body mass (LBM). When comparing those SNPs identified by GWAS with the 226 candidate SNPs used in this thesis (**paper 1**, **2** and **4**), only two SNPs (rs16892496 and rs7832552) within the gene *TRHR* were found in common. Notably, the two *TRHR* SNPs were significantly associated with LBM in the study of Liu et al.[111] However, neither of these SNPs was closely correlated with any muscular parameters in this thesis. Besides nuclear genes, some mitochondrial genes have also been connected with skeletal muscle. Bray et al.[210] summarised 18 mitochondrial genes that have been associated with exercise intolerance, fitness or physical performance. Yet, no mitochondrial genes were analysed in this thesis. Therefore, the incomplete candidate SNP set and the lack of gene expression measurement might contribute to a smaller proportion of explained variance in muscular phenotypes (mass and strength) in a GPS study compared to a heritability study.

Notably, Young[296] summarised in a recent review that even with a complete trait-specific SNP set identified from a GWAS, the problem of missing heritability might still exist. One of the main reasons is that the application of a GWAS was only able

to detect genetic variants that are rather common in the population and have relatively strong connections with a trait, and some very rare genetic variants that are related to a trait might fail to be captured by a GWAS. For example, heritability studies have reported that genetic factors account for more than 68% of the variance in height.[297,298] Yet, Yengo et al.[299] reported that SNPs that were significantly related to height or BMI only explained 24.6% and 6% of the variance, respectively. Yang et al.[300] analysed a large set of common (not height-specific) SNPs (n = 294,831) and could explain 45% of variance in height. Wood et al.[301] conducted a GWAS meta-analysis on height of 253,288 individuals and found that SNPs with a strong association with height explained up to 29% of the variance in height while all common variance together could explain 60% of the variance. These findings indicate that GWASs cannot detect all contributing genetic variations that are related to a trait despite that increasing GWAS sample size will enhance the proportion of explained variance, and consequently, trait-specific SNPs explain less variance than that estimated in a heritability study. Another possible reason is that by estimating heritability, one assumes that the genetic factor is a result of an additive influence from each contributing genotype and therefore fails to take into account any gene-gene[302] and gene-environment interactions.[303] For example, Zuk et al.[302] reported that genetic interactions could account for 80% of the missing heritability in Crohn's disease and suggested that genetic interactions was also important to be examined in estimating heritability.

2.2. Genetic association with *adaptations* in muscle mass and strength after training and detraining

By far, only one study from Thomaes et al.[123] has reported the association between GPS and muscle adaptations to exercise. In the study, 260 coronary artery disease (CAD) patients followed a three-month cardiac rehabilitation program. A data-driven GPS (based on 54 SNPs, selected by backward regression) was calculated to examine the genetic association with exercise-induced changes in fat free mass and knee strength. Their results showed that fat free mass and isometric quadriceps strength were improved by 1.2% (SD 3.6%) and 11.5% (SD 16.0%) after the training. The data-driven GPS explained 2.6% of the variance in isometric quadriceps strength change. In the current work (**paper 1**), exercise induced a 2.7% (SD 7.3%) and 13.3% (SD 15.5%) improvement in muscle mass and isometric knee strength, respectively, while the data-driven GPS explained 27% of the variance in isometric knee strength change. The different results between these two studies might be due

to the difference in training protocols and GPS calculations. In Thomaes' study, the cardiac rehabilitation program mainly consisted of aerobic exercise and calisthenics with a comparatively short intervention period of three months. The training protocol in **paper 1** included both resistance and aerobic training with a comparatively high intensity during a one-year of intervention. Therefore, there was a more intensive and longer-lasting exercise stimulus and a longer genotype*training interaction phase in **paper 1**, resulting (partially) in higher average responses of muscle mass and knee strength. Moreover, Thomaes' study was based on 54 SNPs and only 2 SNPs (*ACVR1B* rs746434 and *AMPD1* rs17602729) were selected as being closely related to isometric knee strength. **Paper 1** selected eight SNPs (none of them were shared with Thomaes' study) that were correlated with isometric knee strength out from a 170-candidate SNP pool. Therefore, the data-driven GPS in **paper** 1 explained more variance in isometric knee strength change after exercise intervention. Notably, the fact that no common SNPs were shared between Thomaes' study and **paper 1** indicates a genetic association with the characteristics (e.g. type, intensity and frequency) of the training intervention. As reported in **paper 1**, the strength-related GPSs were not closely related to aerobic adaptations of the participants. Therefore, further understanding of the susceptibility of SNPs towards different exercise characteristics will be helpful for personalised regimen design in the future.

The genetic association with muscular changes after the cessation of training has never been studied before. If the muscular phenotypes at baseline (**paper 1**) are seen as overall results of environmental and genetic effects over a long (life-) time, then the muscular changes after a one-year cessation of training (termed as "detraining" in **paper 2**) can be roughly regarded as the results of a short-term environmental and genetic interaction (genotype*detraining interaction) characterised by the lack or sudden stop of supervised exercise training. In **paper 2**, a data-driven GPS explained 26–37% of the variance in muscular changes during detraining. This indicates that at least part of the training gains are lost after quitting training depends on your genetic architecture. Since no other studies are available to compare with, limited data on more general longitudinal aging changes can be interpreted. The explained detraining variability in **paper 2** resembled the finding of a 35% genetic contribution to the variance in a 10-year grip strength decline in older twins reported in a heritability study.[59] However, as discussed in **paper 2**, extra data

such as food consumption and physical activity also need to be controlled when analysing the genetic association with muscular changes during detraining.

2.3. Shared genetic variants across *baseline* muscle mass and strength, and muscular phenotypic changes during *training* and *detraining* (based on **paper 1** and **2**)

Paper 1 and **2** explored the genetic association with inter-individual muscular variability at baseline and after exercise among 200 older Flemish adults. **Paper 1** identified seven SNPs (located in six genes) associated with baseline muscle mass and knee strength, 13 SNPs (located in 11 genes) were associated with muscle mass and knee strength changes induced by a one-year of structured training. **Paper 2** identified 46 SNPs (located in 34 genes) that were related to muscular changes after a one year of detraining. In total, 60 SNPs from 43 genes were identified from those two studies. Those identified genes are related to multiple domains such as DNA methylation regulation (*MTHFR*, *MTR* and *MTRR*), growth/differentiation factors (e.g. *IGF1*, *MSTN* and *TGFB1*), hormone receptors (e.g. *ACVR1B*, *ESR1* and *VDR*) and neural factors (e.g. *CNTF*, *RIMS1* and *ZNF804A*). To investigate if a genetic variant is involved at more than one timepoint or condition (baseline, training or detraining), comparisons of data-driven SNPs between different timepoints were made. No common SNP is found among the three timepoints while in total, six SNPs (located in six genes) were shared between each two timepoints (D-Table 3).

The SNP rs3110697 from the gene *IGFBP3* was related to baseline muscle mass, knee strength and knee strength adaptation to training. Based on cultured myoblasts, Foulstone et al.[162] reported that the IGFBP-3 secretion was positively associated with myoblast differentiation. The addition of antisense IGFBP-3 reduced the IGFBP-3 secretion, resulting in decreased skeletal muscle differentiation.[162] The SNP rs3110697 demonstrates a strong connection with plasma IGFBP-3 levels,[163] and is also closely related to the appetite loss in cancer patients, which might further result in weight loss and cachexia.[304] The identification of rs3110697 indicates its possible connection with baseline values of muscle mass and knee strength, and knee strength adaptation to exercise.

Two SNPs (rs10497520 and rs3797297) in *TTN* and *FST* are found in both baseline knee strength and detraining-related strength change, indicating a possible involvement of corresponding SNPs in baseline muscle strength and a muscular

change induced by a one-year detraining period. Titin (encoded by the *TTN* gene) is a crucial protein in striated muscle structure and function. It connects the Z disk to the M line in a sarcomere, and therefore, assists force transmission during muscle contraction and contributes to muscle stiffness in passive stretching. Stebbings et al.[167] reported that the *TTN* rs10497520 C/T polymorphism was associated with skeletal muscle fascicle length. Marathon runners with the T allele had shorter vastus lateralis fascicle length and better performance in a marathon competition.[167] Thomaes et al.[123] studied the rs10497520 polymorphism and reported an association with isokinetic knee extension strength in CAD patients. The *FST* gene encodes follistatin, a protein that is involved in myogenesis regulation. Lee et al.[165] showed that transgenic mice, with high levels of follistatin, exhibited dramatic increases in muscle mass and gastrocnemius muscle fibre size compared to control mice. The researchers further reported that the *FST* mutant mice, with reduced follistatin levels, had significant decrease in muscle size and tetanic force production.[305]

The comparison between training and detraining showed three SNPs (rs2390760, rs3762546 and rs7703033) in the gene *METTL21C*, *MSTN* and *MTRR*, respectively, suggesting that those SNPs might be sensitive to exercise stimulus and loss of the stimulus. METTL21C (encoded by the gene *METTL21C*) is specifically expressed in MYH7-positive skeletal muscle fibers.[306] The gene *METTL21C* was identified as being closely related to bone and muscle function via a bivariate GWAS for paired bone geometry and muscle phenotypes.[178] Cell line analyses further showed that partial silenced *METTL21C* inhibited myoblast differentiation, reduced the amplitude of caffeine-induced peak Ca^{2+} release from the sarcoplasmic reticulum, and promoted cell death.[178] The gene *MSTN* encodes myostatin, a TGF-β family member that negatively regulates skeletal muscle mass. Animal studies have shown that the muscle weights in *MSTN*-knockout mice are almost twice as heavy as that in wild-type mice and such muscle mass difference is primarily due to muscle fibre hypertrophy.[114,307] Schuelke et al.[308] examined a child with a rare *MSTN* mutation (without mature myostatin in serum) and reported that the child showed an extremely larger quadriceps size (7.2 SD above the mean value of age- and sex-matched controls) and a thinner subcutaneous fat layer (2.88 SD lower than the mean value of controls). These findings provide strong evidence of the negative regulator role of *MSTN* in muscle mass in human beings. Exercise is found to trigger an alteration in the myostatin level. Hittel et al.[309] reported that a 6-month moderate

aerobic exercise induced myostatin reduction in both muscle and plasma samples from middle aged adults. The research team also mentioned a notable inter-individual variance in plasma myostatin at pre- and post-training levels, suggesting a possible genetic association with myostatin expression.[309] Methionine synthase reductase (encoded by *MTRR*) plays a crucial role in the metabolic cycle of producing methyl groups for DNA methylation.[183] In rs7703033, the A allele is associated with reduced MTRR expression, leading to decreased DNA methylation level. The identification of the *MTRR* gene in both training and detraining timepoints indicates a possible involvement of differential DNA methylation in individual muscular responses to exercise stimulus.

2.4. DNA methylation and sarcopenia

Previous studies compared DNA methylation differences between older and young adults based on various tissues such as saliva,[100] blood,[229,230] muscle,[101] skin[231] and brain[232], in which age might be an important factor to induce methylation changes.[98] To restrict the age effect on DNA methylation and to better explore sarcopenia-related methylation differences per se, we controlled the ageing effect by selecting generally age-matched older women with a limited age difference (**paper 3** and **4**). Possibly also because of the control for age, the age (as a covariate) was not significantly related to most muscular phenotypes (vastus lateralis size, elbow and knee strength) in **paper 4**. Moreover, to better explore the association between DNA methylation and skeletal muscle, methylation profiles were compared between two groups with significant muscle differences (sarcopenic vs. non-sarcopenic). Therefore, the observed methylation differences might be largely related to inter-individual variability in muscular phenotypes among older adults with similar ages.

This thesis (**paper 3**) analysed DNA methylation differences related to sarcopenia. To our best knowledge, no similar studies have been reported before. As summarised in **paper 3**, a general hypermethylation in gene promoters and hypomethylation in gene bodies was correlated with sarcopenia. In total, 6,258 sarcopenia-related differentially methylated CpGs (dmCpGs, $p < 0.01$) were identified. Genes containing these dmCpGs were involved in pathways of multiple domains such as muscle function, actin cytoskeleton regulation and energy metabolism. Seven genes (*HSPB1*, *PBX4*, *CNKSR3*, *ORMDL3*, *MIR10A*, *ZNF619* and *CRADD*) were found with the same methylation direction as that reported in previous ageing-related methylation studies based on blood samples,[229,230]

indicating a possible connection between these seven genes and ageing-related muscle degeneration. Notably, although **paper 3**, which was based on blood tissue, also shared multiple CpGs with methylation studies on ageing[101] and resistance training in young adults,[102,104] which were based on muscle tissue, the methylation directions of those CpGs were not always consistent between the two tissues. The inter-tissue inconsistency of methylation supports the observation that DNA methylation pattern is tissue specific,[95] and therefore, it might not be feasible to estimate the methylation status in skeletal muscle based on the methylation data obtained from blood tissue.

Instead of selecting sarcopenia-related CpGs from dmCpGs (n = 6,258), which was identified in **paper 3** by comparing M values between the sarcopenic and non-sarcopenic group, **paper 4** conducted the selection based on all initially measured CpGs (n = 788,074). The rationale includes (1) to select (possibly) more CpGs that might be related to sarcopenia status without the restriction of p-values from an association test (t-test in **paper 3**), (2) to control for the collinearity of selected CpGs, (3) to obtain the weight of each CpG site for the calculation of a sarcopenia-driven methylation score. The concern behind the first rationale is that the dmCpG set (in **paper 3**) was determined by an arbitrary p-value threshold of 0.01, therefore, some CpGs which were closely correlated with sarcopenia might be excluded due to a higher p-value (greater than 0.01). While the lasso regression analysis on the initial CpG dataset allows to choose CpGs without the p-value restriction, therefore, many representative CpGs that are related to sarcopenia can be selected. The concern behind the second and third reasons are related to the research aim of each paper. **Paper 3** explored the DNA methylation difference and corresponding pathways in association with sarcopenia, therefore, the main purpose was to identify the CpGs that showed significant methylation differences between the sarcopenic and non-sarcopenic group. Because of this, a p-value threshold was used for CpG selection. In this approach, CpGs that locate closely in the same gene region or are functionally correlated might be identified together. **Paper 4** aimed to investigate the association between DNA methylation profiles (represented by a methylation score) and muscular phenotypes. Therefore, in the calculation of a sarcopenia-driven methylation score, the collinearity between CpG sites should be controlled and the weight of each CpG needs to be estimated. For this purpose, a lasso regression was used to select CpGs that had relatively strong correlations (non-zero coefficients) with sarcopenia, and were not closely related to each other. As

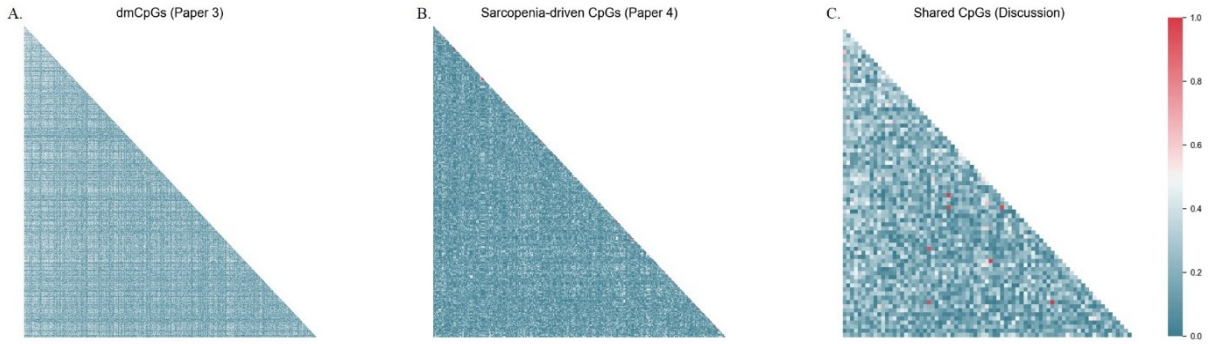
demonstrated in D-Table 4 and D-Figure 1, the proportion of weakly correlated CpGs selected by the lasso regression (**paper 4**) is higher than that in dmCpGs identified in **paper 3**.

Yet, stricter CpG selection criteria can still be applied by combining both the lasso regression results (n = 379) and the p-value threshold (n = 6,258), which gives a more representative set of sarcopenia-related CpGs (n = 76, D-Supplementary Table 1A). The shared CpGs have the smallest count of strong correlation, and the percentages of weak and moderate correlations in shared CpGs are between that in dmCpGs and sarcopenia CpGs (D-Table 4, D-Fig 1). Chi-squared test showed that the distributions of weak, moderate and strong correlations are different among the three groups ($p < 0.01$).

D-Table 4. Description of correlations between CpGs identified in paper 3, 4 and the current discussion.

Correlation	dmCpGs (paper 3)		Sarcopenia-driven CpGs (paper 4)		Shared CpGs (current discussion)	
	Count	Percentage	Count	Percentage	Count	Percentage
Weak (< 0.3)	13,726,712	70.11	66,027	92.18	2,230	78.25
Moderate (0.3 - 0.7)	5,846,997	29.86	5,542	7.74	613	21.51
Strong (> 0.7)	4,444	0.03	62	0.08	7	0.24

The count represents the total correlation count among CpGs: n(n-1)/2, in which n is the amount of CpGs.

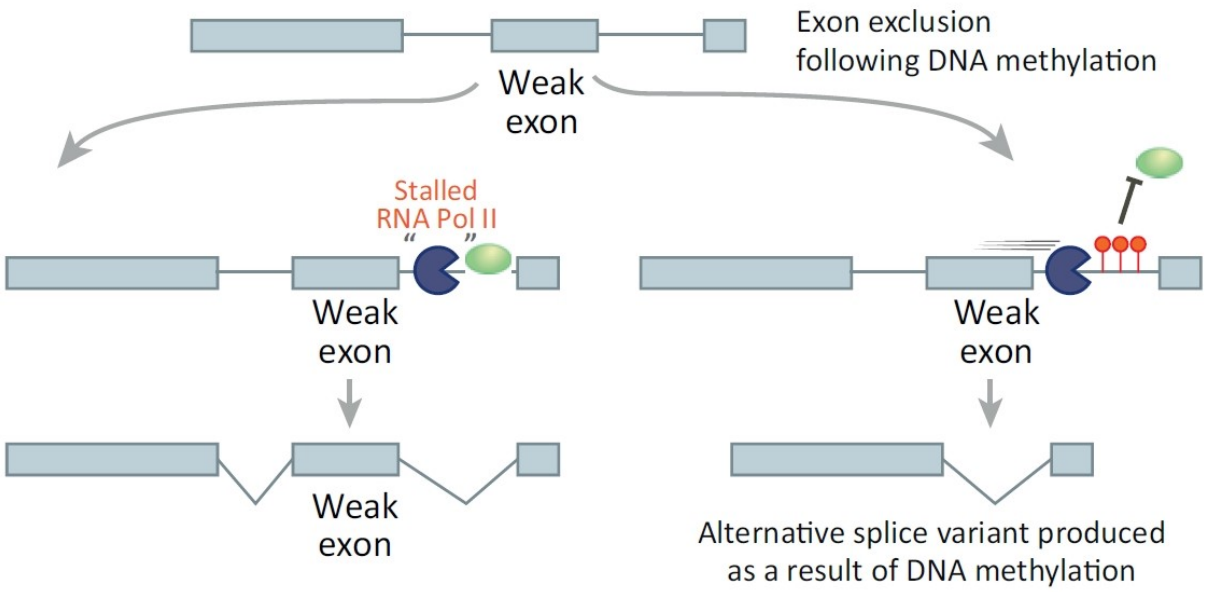


*D-Figure 1. Correlation heatmap of CpGs. (A) dmCpGs (n = 6,258) identified by p-value (p < 0.01) in **paper 3**, (B) sarcopenia-driven CpGs (n = 379) identified by the lasso logistic regression in **paper 4**, (C) shared CpGs (n = 76) in this discussion.*

Gene ontology (GO) enrichment analysis on these 76 CpGs identified 223 terms that are significantly enriched (unadjusted p-value < 0.05) (D-Supplementary Table

1A). The top 20 GO terms are associated with many muscle-related activities such as neural control, signal transduction, calcium ion transport and myoblast differentiation (D-Supplementary Table 1B). The “synapse assembly” is the most significantly enriched GO term which includes three CpGs (cg25340050, cg15714846 and cg19524037) located in different regions (1stExon, 3'UTR and gene body) of the genes *PCDHB16*, *FZD5* and *NRG1* (D-Supplementary Table 1C).

Multiple studies have analysed the association between exon methylation and gene expression.[310–313] Brenet et al.[311] analysed genome-wide DNA methylation with gene expression and found a close negative association between the first exon methylation level and DNA transcription. A recent study by Shayevitch et al.[313] showed that the methylation of exon regions helped to regulate alternative splicing (AS, D-Figure 2), a mechanism that contributes to transcriptomic and proteomic diversity by generating multiple mRNA products from a single gene.[310] The research team reported that the lack of DNA methylation inhibited the infusion of intragenic exons during the formation of mature mRNA.[313] The same results were also found by Li et al.,[312] who showed that gene expression was negatively correlated with methylation levels in promoters and first exons, while exon expression was positively associated with methylation densities in intragenic exons. The result of hypomethylated cg25340050 (lower methylation level in the sarcopenic group than non-sasrcopenic group, D-Supplementary Table 1C) located in the first exon of *PCDHB16* indicated a down-regulated PCDHB16 expression in the sarcopenic group, and the *PCDHB16* gene is relate to the function of cell-cell neural connection.[314]



D-Figure 2. RNA alternative splicing regulated by intragenic DNA methylation. Without DNA methylation, exons will be kept in mature mRNA. With DNA methylation, some exons will be dropped in the formation of mature mRNA. Adapted from Baker-Andresen, 2013.[315]

The gene *FZD5* encodes frizzled class receptor 5, a protein that is regarded as the receptor for Wnt5A ligand.[314] Wnt5A is involved in multiple cellular processes such as the development of the reproductive tract, the inducement of gland formation and the process of oestrogen mediated cellular and molecular responses in uterine tissue.[316] The identification of gene *FZD5* suggests possible cell development and molecular response changes associated with sarcopenia. The gene *NRG1* encodes neuregulin 1, a glycoprotein which plays a role in synaptic plasticity. Based on a study in mice, Agarwal et al.[317] demonstrated that abnormal (deficient or overexpressed) expression of Neuregulin 1 led to disrupted hippocampal plasticity (with impaired long-term potentiation), and imbalanced excitatory and inhibitory neurotransmission. As a result, the learning ability and memory might also be disrupted.[318] Since DNA methylation of the gene body is associated with increased gene expression,[94,241] the hypomethylated cg19524037 located in the gene body of *NRG1* (D-Supplementary Table 1C) might indicate a lower expression of Neuregulin 1 in association with sarcopenia. However, further gene expression analysis in blood samples would be needed to confirm this assumption.

KEGG pathway analysis on the 76 CpGs showed that only the pathway “synthesis and degradation of ketone bodies” is significantly enriched (unadjusted p-value = 0.02, D-Supplementary Table 1D). Ketone bodies, mainly acetoacetate (AcAc) and β-hydroxybutyrate (β-HB), are generated from fatty acid oxidation products in the liver under a glucose starvation condition.[319] They serve as an alternative fuel source for peripheral tissues such as brain and skeletal muscle.[319] After being generated, the ketone bodies diffuse into the bloodstream. In extrahepatic tissues, Ketone bodies are converted into acetyl-CoA with the catalysis by 3-oxoacid CoA-transferase (encoded by the *OXCT1* gene).[320] The acetyl-CoA is later transported into mitochondria for energy production. The pathway “synthesis and degradation of ketone bodies” (10 genes in the pathway) contains one hypermethylated (higher methylation level in the sarcopenic group than non-sarcopenic group) CpG in the *OXCT1* gene (D-Supplementary Table 1E), implying a possible alteration in energy production accompanied with sarcopenia.

Notably, the “mTOR signalling pathway” (152 genes in the pathway) shows a marginal significance (unadjusted p-value = 0.05, D-Supplementary Table 1D). The mTOR signalling pathway is known for promoting muscle growth.[321] The animal study conducted by Bodine et al.[261] showed that the mTOR pathway was upregulated during muscle hypertrophy and downregulated during atrophy. Moreover, with the presence of mTOR blocker, muscle hypertrophy was also inhibited.[261] Studies on human skeletal muscle have shown that acute resistance training increases the phosphorylation of ribosomal protein S6, a substrate of p70 S6 kinase ($p70^{S6k}$) in the mTOR pathway[322], and leads to rapid translocation of mTOR/LAMP2 towards the cell membrane with concurrent increase in mRNA translation capacity.[323] These changes in the mTOR pathway might partly explain the increased muscle protein turnover[324] and enhanced myofibrillar protein synthesis[325] after acute resistance training. The KEGG analysis identified two genes (*PRR5* and *FZD5*) from the mTOR pathway. The function of *FZD5* has been previously discussed. The gene *PRR5* encodes proline rich 5 protein, which is a component of the mTOR complex 2 (mTORC2). Woo et al.[326] demonstrated that *PRR5* silencing reduced the expression of the growth factor receptor PDGFR, and repressed Akt and S6K1 phosphorylation through the PDGF signalling pathway. The discovery of the gene *PRR5* and *FZD5* suggests a change of muscle growth ability in sarcopenic women compared with non-sarcopenic women.

D-Table 5. Genetic and methylation scores in **paper 4**.

Genetic and methylation score	Calculation method	Rationale
Muscle-driven genetic predisposition score (GPS_{SNP})	A summed GPS of seven muscle-related SNPs	This score is used to represent an **overall** genetic effect based on DNA sequence variants in **seven** genes that are established for their associations with skeletal muscle phenotypes.
Sarcopenia-driven methylation score (MS_{SAR})	A weighted sum of the selected CpG methylation levels (the weight for each CpG site is the coefficient from the lasso regression)	This score is used to represent an **overall** sarcopenia-related methylation level.
SNP-driven methylation score (MS_{SNP})	An average of the methylation levels in muscle-related genes	This score is used to represent an **average** methylation level **across selected** muscle-related genes where the seven muscle-related SNPs locate. Together with GPS_{SNP}, it is designed to evaluate associations between seven selected muscle-related genes at the genetic sequence level (GPS_{SNP}) and the methylation level (MS_{SNP}) with individual variation in muscular phenotypes.

Gene-wise combined score	The ratio of a SNP score to the average methylation level in promoters of the corresponding gene (Seven gene-wise combined scores)	This score is used to represent a genetic sequence and methylation level **interaction within each** muscle-related gene.

2.5. DNA methylation profile scores and inter-individual variability in muscle morphology and strength

Regardless of the methods used for sarcopenia-related CpGs selection (p-value based in **paper 3**, lasso regression based in **paper 4**, or combined criteria discussed previously in section 2.4), the identified CpGs might be closely related to sarcopenia status as well as inter-individual variability in muscle morphology and strength.

Paper 4 is the first study to analyse both genetic and methylation profile scores and their predictive value in ageing muscle. In **paper 4**, multiple models were built based on separate/combined genetic and methylation scores at an overall/gene-specific level (D-Table 5). In short we found that genetic profile scores based on seven selected genes (GPS_{SNP}) explained up to 4.4% of the variance in muscle size and strength, methylation levels in the seven selected genes (MS_{SNP}) explained up to 8% in the studied phenotypes, while the methylation profile score based on sarcopenia-related CpGs (MS_{SAR}) explained 10.1–40.1% of the individual differences in muscle morphology and strength in our sample of older women. These results showed that genetic and methylation profiles on several representative genes were not strong enough to fully explain muscular variability. Moreover, by comparing the Akaike information criterion (AIC), the model with only MS_{SAR} showed the best performance in explaining the variance in muscle size and strength. This, again, indicates that using the data from a small set of representative genes cannot well explain muscular variability. Notably, as discussed in section 2.1, even the data-driven GPS based on a larger gene set (**paper 1)** only explained up to 3.2% of the variance in muscle mass and strength. Therefore, it is possible that the genetic architecture only accounts for a small portion of muscular variability in ageing muscle – or we have not yet used the optimal methodology to include all contributing genetic factors, while a larger proportion of the variance is taken up by DNA methylation. Since DNA methylation is representing the sum of short-term and long-term environmental factors, the finding that methylation levels explains a larger proportion of the variance in muscle morphology and strength than genetic profiles might indicate that environmental elements account for more variance than genetic factors in skeletal muscle during ageing, which is supported by several heritability studies.[59,61,289]

To further explore if there is a possible synchronization between the SNP genotype and the methylation status of a corresponding gene, i.e. the gene with a higher SNP score has a lower promoter methylation level (a higher expression level), ANOVA

analysis was made to compare gene promoter methylation levels between SNP genotype groups in each gene. Although no significant difference was detected, the results showed that SNP scores in the gene *ACE*, *CNTF*, *FTO* and *HIF1A* were in the same direction as that in promoter methylation levels (D-Table 6). Participants with a higher SNP score also had a higher promoter methylation level in the corresponding gene. Therefore, this thesis shows that genetic structures do not necessarily cooperate with the methylation status of corresponding genes. Interestingly, if the scoring of muscle-predisposing alleles is based on dominant alleles that are closely associated with increased gene expression in skeletal muscle (based on the GTEx database),[292] as in the case of SNP scores in the gene *CNTF* and *FTO*, we observed an opposite direction of promoter methylation levels compared to SNP scores in these two genes. Therefore, higher SNP scores in the gene *CNTF* and *FTO* tend to be associated with lower promoter methylation levels and higher gene expression levels.

As a pilot study, **paper 4** indicates that the methylation levels of sarcopenia-related CpGs can be possibly used to identify older (female) adults who are susceptable to muscle degeneration. Based on this identification, many preventive interventions, such as a personalised exercise regimen, extra nutrient supplements and additional home care service, can be applied before the occurrence of functional limitations caused by degenerated skeletal muscle. However, at the current stage, multiple validations on different independent cohorts are still needed to verify the predictive power of these CpG sites.

D-Table 6. SNP scores and methylation levels of promoter regions in corresponding genes from **paper 4**.

SNP#		Mean	SD	Dominant allele in GTEx*
ACTN3 (rs1815739 T/**C**)				-
	0	0.3947	0.0049	
	1	0.3972	0.0101	
	2	0.3966	0.0129	
ACE (rs4341 C/**G**)				-
	0	0.4487	0.0071	

	1	0.4489	0.0066	
	2	0.451	0.0077	
CNTF (rs1800169 A/**G**)				A
	1	0.8422	0.0105	
	2	0.8474	0.0155	
FTO (rs9939609 A/**T**)				A
	0	0.2446	0.0048	
	1	0.2451	0.003	
	2	0.2469	0.0047	
HIF1A (rs11549465 C/**T**)				T
	0	0.1811	0.0059	
	1	0.1836	0.0057	
VDR (rs2228570 G/**A**)				-
	0	0.2238	0.0018	
	1	0.2231	0.0036	
	2	0.2225	0.0034	

SNPs examined in **paper 4**. Alleles predisposing to muscle mass and strength (based on published association studies) are presented in bold. The *MSTN* SNP (rs1805086) is not included because all the participants in **paper 4** had the same TT genotype (T allele frequency in general population: 96.9%).

* The dominant allele is the allele associated with increased gene expression in skeletal muscle based on the GTEx database. Alleles that are not significantly related to gene expression in skeletal muscle are not presented.

2.6. Summary of SNPs and pathways (methylation based) associated with inter-individual variance in muscle mass and strength

This thesis identifies 43 genes (**paper 1** and **2**) that are associated with individual variability of muscle mass and strength under different conditions (baseline, exercise training and detraining) among older people. Moreover, based on sarcopenia-related CpGs, 37 pathways are identified as being related to muscular differences (**paper 3**). These genes and pathways are involved in multiple biological

processes such as hormonal function (e.g. *IGF1*, *IGF2*, insulin resistance, thyroid hormone signaling pathway), muscle function (e.g. *MSTN*, *TTN*, apelin signaling, cGMP-PKG signaling), actin cytoskeleton regulation (e.g. phosphatidylinositol signaling, focal adhesion, adherens junction), energy metabolism (e.g. thermogenesis, AMPK signaling, glucagon signaling), neural control (e.g. *CNTF*, *RIMS1*, *ZNF804A*, axon guidance, GABAergic synapse), signal transduction (e.g. Wnt signaling, MAPK signaling, cAMP signaling), methylation regulation(e.g. *MTHFR, BHMT, MTR*) and cell regeneration (e.g. *CCL2*, *CCR2*, cell cycle, oxytocin signaling). All these processes are connected directly/indirectly to muscle mass/strength and can probably (partially) explain inter-individual variations in muscular phenotypes. Appendix table 1 has categorised these genes and pathways in terms of muscle mass, muscle strength, muscle recovery and methylation regulation. 13 genes and 10 pathways are connected with muscle mass, 23 genes and 18 pathways are related to muscle strength, 9 pathways are related to both muscle mass and strength, 2 genes are associated with muscle recovery and 5 genes are connected to methylation regulation.

3. Limitations and suggestions for future research

(1) Studies in this thesis have limited sample sizes, which affect statistical power. Although the sample size in the one-year exercise training study is large given the nature and time-demanding aspects of a supervised intervention study of long duration, it is still rather small for a genetic association study.[327] Therefore, additional studies with a large sample size are needed to validate our findings. Within the MMU study in older women, both the thresholds used to define sarcopenia, as well as limited financial resources resulted in a small set of women for which an Infinium MethylationEPIC BeadChip methylation analysis could be performed. Moreover, the participants in the methylation studies of this thesis are older women. Compared to older men, older women have lower sex hormone levels and less relative loss in muscle mass and strength during muscle degeneration.[328] Other hormonal functions, such as the insulin and glucagon responses for glucose homeostasis,[329] and thyroid function,[330] are also reported as sex-specific. In this thesis, methylation level differences were identified in genes contributing to multiple hormone-related pathways, e.g. insulin resistance, glucagon signaling pathway and thyroid hormone signaling pathway (Appendix table 1). Therefore, the methylation findings in **paper 3** and **4** might not be fully generalised to older men. In the future,

similar methylation researches on sarcopenia-related DNA methylation changes **in older men** are needed.

Since no CpGs were found with significance after the false discovery rate (FDR) correction at 5%, an arbitrary p-value threshold ($p < 0.01$) was used in **paper 3**. The selection of a p-value threshold can be a weakness in this thesis because the selected p-value will affect the identification of dmCpGs and subsequent GO and KEGG analysis results. In ageing-related methylation studies, Heyn et al.[229] analysed blood methylation patterns between a newborn and a centenarian, and identified 214 dmCpGs after FDR correction at 1%. Bell et al.[230] analysed DNA methylation (blood based) of 172 female twins (aged between 32 and 80 yrs) and identified 490 hypermethylated dmCpGs related to chronological age after epigenome-wide association scans (EWAS) and FDR correction at 5%. In future studies, **a larger sample size** will be helpful to identify significant sarcopenia-related dmCpGs after FDR correction.

(2) The methylation analysis in this thesis is based on DNA methylation data of blood cells. Although methylation profiles were normalised for white blood cell counts, which were estimated through an R package "FlowSorted.Blood.EPIC", the methylation profiles would be more accurate if **flow cytometry measurements of cell composition** had been made from blood samples. Additionally, since methylation is tissue-specific, future methylation studies **on muscle biopsies** in older subjects can be more informative for the understanding of muscle degeneration and the derived methylation score might explain more inter-individual muscular variance.

(3) On the other hand, despite that DNA methylation patterns are tissue-specific, future **comparisons of DNA methylation differences based on blood and skeletal muscle tissues** are still needed to identify CpGs with a fixed (either positive or negative) correlation between both tissues. If some overlapping CpGs can be found, then the more easily accessible blood sample can be used in further studies and applications compared to muscle biopsy.

(4) This thesis only used stepwise and lasso regression for SNP and CpG selection. In fact, **many other techniques** can be used in the selection of genetic and methylation markers, as well as the calculation of GPS and methylation profile scores.

For instance, instead of using a preselected candidate gene set, a GWAS can be used to select muscle-related genetic variants. Shah et al.[130] used a GWAS and a methylome-wide association study (MWAS) to identify SNPs and CpGs that were significantly associated with BMI/height, and tested the BMI/height models with these significant genetic and methylation markers in other independent cohorts.

The lasso regression used in this thesis is one of the machine learning methods for variable selection. The random forest (RF) method has also been used for variable selection in previous genetic studies. Dogan et al.[131] used the RF method with combined genetic and methylation markers for coronary heart disease (CHD) detection. The research team first selected SNPs with low linkage disequilibrium (threshold of 0.5) and close correlation (p-value < 0.1) with CHD status. They also selected CpGs based on their correlations with CHD status (more than 0.1) and the correlations between each other (less than 0.8). Subsequently, RF models were built based on selected genetic and methylation markers for a further identification of SNPs and CpGs that contribute to a high prediction for CHD. In the end, the predictive model with the identified SNPs and CpGs were evaluated in an independent test set with an accuracy higher than the conventional classification (78% vs. 65%). With a large dataset, similar methods can also be used in the identification of sarcopenia.

Additionally, different methodologies of processing SNP data can be applied. In this thesis, the SNP is scored as 0, 1, and 2 based on the number of predisposing alleles. Such process is based on the assumption that an allele, which is not favourable to muscle, does not contribute to any muscular phenotype. If assuming that an allele that is less favourable to muscle also contributes to muscular phenotypes, then a different scoring system of 1, 1.5, and 2 can be applied. Additionally, we should be aware that the SNP scoring used in this thesis is different from the conventional idea of the “risk allele” in disease studies, in which the weights of SNP genotypes are not given arbitrarily, but are based on the effect size of the genotype in a population. Moreover, besides a scoring system, the genotype in a SNP can also be converted into two dummy variables in further analysis.

(5) In this thesis, genetic structure and DNA methylation explains up to 43% of the variance in muscle mass and strength, which is close to the genetic influence (31–67%) on muscle mass and knee strength estimated in heritability studies (as described in section 2.1). This also means that 57% of inter-individual variability in

muscle size and strength characteristics are related to other variables that are not included in the analysis. Therefore, future studies can be conducted with the inclusion of **additional genetic factors** (e.g. muscle-related mitochondrial gene variants, GWAS identified gene variants, copy number variation, gene interactions), **gene expression data**, **epigenetic factors** (e.g. histone modification, noncoding RNA), and measured **environmental factors** (e.g. food pattern, physical activity level, living habits and toxic environmental factors).

(6) The methylation analysis in this thesis is cross-sectional, therefore, only the inter-individual variability in muscle morphology and strength is studied. In the future, a **longitudinal study of methylation changes** during the ageing process can also enhance the understanding of ageing. Moreover, methylation markers can be identified that predict muscular changes during ageing process, which will be helpful for preventive screening for older adults who have a faster muscle degeneration speed so that preventive treatments (e.g. exercise training and nutrient supplementation) can be made. Future intervention studies in older adults can also investigate whether the methylation status at the set of identified dmCpGs (**paper 3**) or sarcopenia-based CpGs (**paper 4**) can be altered towards the more 'healthy' pattern following strength or other types of training.

4. General conclusion

The current thesis contributes to the understanding of the roles that genetic sequence variation and DNA methylation play in individual differences in muscle morphology and strength in older people.

Genetic architecture – as captured in a genetic predisposition score - is not only related to one's strength level in the untrained state, but also partially determines whether you are a high- or low responder after one year of training as well as after a period of detraining. Specifically, the results show that genetic predisposition scores account for the variance in muscle mass and strength by 0.7–3.2% at baseline, 14–27% after a one-year training intervention, and 26–37% after a one-year cessation of training. Additional genetic information is needed to improve the predictive power of these genetic predisposition scores.

Environmental factors like nutrition, physical activity or other lifestyle choices can affect the studied phenotypes through epigenetic mechanisms, of which (de-)methylation of gene regions is one mechanism influencing gene expression. This thesis further provided new insights on the role of DNA methylation by

identifying sarcopenia-related methylation differences in the DNA of blood cells in older women. A general hypermethylation in gene promoters and hypomethylation in gene bodies was associated with sarcopenia. Genes that were identified as differentially methylated are mainly connected with muscle function, actin cytoskeleton regulation and energy metabolism. The construction of methylation scores showed that a sarcopenia-driven methylation score explained 10.1–40.1% of the variance in muscle size and strength at baseline level in older women. Although further replication of these findings is needed, our results indicate that the use of a sarcopenia-driven methylation score to estimate muscle condition and sarcopenia-risk is merited.

References

1. Faulkner JA, Larkin LM, Claflin DR, Brooks S V. Age-related changes in the structure and function of skeletal muscles. *Clin Exp Pharmacol Physiol* 2007;**34**:1091–1096.

2. von Haehling S, Morley JE, Anker SD. An overview of sarcopenia: Facts and numbers on prevalence and clinical impact. *J Cachexia Sarcopenia Muscle* 2010;**1**:129–133.

3. Frontera W, Hughes VA, Fielding RA, Fiatarone M, Evans W, Roubenoff R. Aging of Skeletal Muscle: A 12-yr longitudinal study. *J Appl Physiol J Gerontol A Biol Sci Med Sci J Appl Physiol* 2000;**88**:1321–1326.

4. Goodpaster BH, Park SW, Harris TB, Kritchevsky SB, Nevitt M, Schwartz A V. *et al.* The loss of skeletal muscle strength, mass, and quality in older adults: The Health, Aging and Body Composition Study. *Journals Gerontol - Ser A Biol Sci Med Sci* 2006;**61**:1059–1064.

5. Charlier R, Mertens E, Lefevre J, Thomis M. Muscle mass and muscle function over the adult life span: A cross-sectional study in Flemish adults. *Arch Gerontol Geriatr* 2015;**61**:161–167.

6. Cruz-Jentoft AJ, Bahat G, Bauer J, Boirie Y, Bruyère O, Cederholm T *et al.* Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing* 2018;**39**:412–23.

7. Schaap LA, Koster A, Visser M. Adiposity, Muscle Mass, and Muscle Strength in Relation to Functional Decline in Older Persons. *Epidemiol Rev* 2013;**35**:51–65.

8. Schaap LA, van Schoor NM, Lips P, Visser M. Associations of Sarcopenia Definitions, and Their Components, With the Incidence of Recurrent Falling and Fractures: The Longitudinal Aging Study Amsterdam. *Journals Gerontol Ser A* 2018;**73**:1199–1204.

9. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, Boirie Y, Cederholm T, Landi F *et al.* Sarcopenia: European consensus on definition and diagnosis. *Age Ageing* 2010;**39**:412–423.

10. Baumgartner RN, Koehler KM, Gallagher D, Romero L, Heymsfield SB, Ross RR *et al.* Epidemiology of Sarcopenia among the Elderly in New

Mexico. *Am J Epidemiol* 1998;**147**:755–763.

11. Janssen I, Heymsfield SB, Ross R. Low Relative Skeletal Muscle Mass (Sarcopenia) in Older Persons Is Associated with Functional Impairment and Physical Disability. *J Am Geriatr Soc* 2002;**50**:889–896.

12. Chien M-Y, Huang T-Y, Wu Y-T. Prevalence of Sarcopenia Estimated Using a Bioelectrical Impedance Analysis Prediction Equation in Community-Dwelling Elderly People in Taiwan. *J Am Geriatr Soc* 2008;**56**:1710–1715.

13. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J *et al.* Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001;**56**:M146-56.

14. Newman AB, Kupelian V, Visser M, Simonsick E, Goodpaster B, Nevitt M *et al.* Sarcopenia: Alternative Definitions and Associations with Lower Extremity Function. *J Am Geriatr Soc* 2003;**51**:1602–1609.

15. Delmonico MJ, Harris TB, Lee J-S, Visser M, Nevitt M, Kritchevsky SB *et al.* Alternative Definitions of Sarcopenia, Lower Extremity Performance, and Functional Impairment with Aging in Older Men and Women. *J Am Geriatr Soc* 2007;**55**:769–774.

16. Janssen I, Baumgartner RN, Ross R, Rosenberg IH, Roubenoff R. Skeletal muscle cutpoints associated with elevated physical disability risk in older men and women. *Am J Epidemiol* 2004;**159**:413–21.

17. Su Y, Hirayama K, Han T, Izutsu M, Yuki M. Sarcopenia Prevalence and Risk Factors among Japanese Community Dwelling Older Adults Living in a Snow-Covered City According to EWGSOP2. *J Clin Med* 2019;**8**:291.

18. Zengin A, Jarjou LM, Prentice A, Cooper C, Ebeling PR, Ward KA. The prevalence of sarcopenia and relationships between muscle and bone in ageing West-African Gambian men and women. *J Cachexia Sarcopenia Muscle* 2018;**9**:920–928.

19. Shafiee G, Keshtkar A, Soltani A, Ahadi Z, Larijani B, Heshmat R. Prevalence of sarcopenia in the world: A systematic review and meta-analysis of general population studies. *J Diabetes Metab Disord* 2017;**16**:1–10.

20. UN Population Division. Population Ageing and Sustainable Development.

Popul Facts 2017 doi:ST/ESA/SER.A/290.

21. Visser M, Goodpaster BH, Kritchevsky SB, Newman AB, Nevitt M, Rubin SM *et al.* Muscle mass, muscle strength, and muscle fat infiltration as predictors of incident mobility limitations in well-functioning older persons. *Journals Gerontol - Ser A Biol Sci Med Sci* 2005;**60**:324–333.

22. Landi F, Liperoti R, Russo A, Giovannini S, Tosato M, Capoluongo E *et al.* Sarcopenia as a risk factor for falls in elderly individuals: Results from the ilSIRENTE study. *Clin Nutr* 2012;**31**:652–658.

23. Newman AB, Kupelian V, Visser M, Simonsick EM, Goodpaster BH, Kritchevsky SB *et al.* Strength, but not muscle mass, is associated with mortality in the health, aging and body composition study cohort. *J Gerontol Ser A Biol Sci Med Sci* 2006;**61**:72–77.

24. Sousa AS, Guerra RS, Fonseca I, Pichel F, Ferreira S, Amaral TF. Financial impact of sarcopenia on hospitalization costs. *Eur J Clin Nutr* 2016;**70**:1046–1051.

25. Kyle UG, Genton L, Slosman DO, Pichard C. Fat-free and fat mass percentiles in 5225 healthy subjects aged 15 to 98 years. *Nutrition* 2001;**17**:534–541.

26. Gallagher D, Ruts E, Visser M, Heshka S, Baumgartner RN, Wang J *et al.* Weight stability masks sarcopenia in elderly men and women. *Am J Physiol - Endocrinol Metab* 2000;**279**:366–375.

27. Janssen I, Heymsfield SB, Wang Z, Ross R. Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. *J Appl Physiol* 2000;**89**:81–88.

28. Maltais ML, Desroches J, Dionne IJ. Changes in muscle mass and strength after menopause. *J Musculoskelet Neuronal Interact* 2009;**9**:186–197.

29. Sipilä S, Narici M, Kjaer M, Pöllänen E, Atkinson RA, Hansen M *et al.* Sex hormones and skeletal muscle weakness. *Biogerontology* 2013;**14**:231–245.

30. Haizlip KM, Harrison BC, Leinwand LA. Sex-Based Differences in Skeletal Muscle Kinetics and Fiber-Type Composition. *Physiology* 2015;**30**:30–39.

31. Bea JW, Zhao Q, Cauley JA, Lacroix AZ, Bassford T, Lewis CE *et al.* Effect

of hormone therapy on lean body mass, falls, and fractures: 6-year results from the Women's Health Initiative hormone trials. *Menopause* 2011;**18**:44–52.

32. Rowan SL, Rygiel K, Purves-Smith FM, Solbak NM, Turnbull DM, Hepple RT. Denervation causes fiber atrophy and myosin heavy chain co-expression in senescent skeletal muscle. *PLoS One* 2012;**7**.

33. Aare S, Spendiff S, Vuda M, Elkrief D, Perez A, Wu Q *et al.* Failed reinnervation in aging skeletal muscle. *Skelet Muscle* 2016;**6**:1–13.

34. Lexell J, Taylor CC, Sjöström M. What is the cause of the ageing atrophy? Total number, size and proportion of different fiber types studied in whole vastus lateralis muscle from 15- to 83-year-old men. *J Neurol Sci* 1988;**84**:275–294.

35. Miljkovic N, Lim J-Y, Miljkovic I, Frontera WR. Aging of Skeletal Muscle Fibers. *Ann Rehabil Med* 2015;**39**:155.

36. Malafarina V, Úriz-Otano F, Iniesta R, Gil-Guerrero L. Sarcopenia in the elderly: Diagnosis and treatment. *Maturitas* 2012;**71**:109–114.

37. International Working Group on Sarcopenia. Sarcopenia: An Undiagnosed Condition in Older Adults. Current Consensus Definition: Prevalence, Etiology, and Consequences. International Working Group on Sarcopenia. *J Am Med Dir Assoc* 2011;**12**:249–256.

38. Rooyackers OE, Adey DB, Ades PA, Nair KS. Effect of age on in vivo rates of mitochondrial protein synthesis in human skeletal muscle. *Proc Natl Acad Sci U S A* 1996;**93**:15364–9.

39. Porter C, Hurren NM, Cotter M V., Bhattarai N, Reidy PT, Dillon EL *et al.* Mitochondrial respiratory capacity and coupling control decline with age in human skeletal muscle. *Am J Physiol - Endocrinol Metab* 2015;**309**:E224–E232.

40. Proctor DN, Balagopal P, Nair KS. Age-Related Sarcopenia in Humans Is Associated with Reduced Synthetic Rates of Specific Muscle Proteins. *J Nutr* 1998;**128**:351S-355S.

41. Gajdosik RL, Vander Linden DW, Williams AK. Influence of age on length and passive elastic stiffness characteristics of the calf muscle-tendon unit of

women. *Phys Ther* 1999;**79**:827–38.

42. Peterson MD, Rhea MR, Sen A, Gordon PM. Resistance exercise for muscular strength in older adults: A meta-analysis. *Ageing Res Rev* 2010;**9**:226–237.

43. Ferrari R, Fuchs SC, Kruel LFM, Cadore EL, Alberton CL, Pinto RS *et al.* Effects of different concurrent resistance and aerobic training frequencies on muscle power and muscle quality in trained elderly men: a randomized clinical trial. *Aging Dis* 2016;**7**:697–704.

44. Villareal D, Aguirre L, Gurney AB, Waters D. Aerobic or resistance exercise, or both, in dieting obese older adults. *N Engl J Med* 2017;**376**:1943–1955.

45. Cardinale M, Bosco C. The use of vibration as an exercise intervention. *Exerc Sport Sci Rev* 2003;**31**:3–7.

46. Delecluse C, Roelants M, Verschueren S. Strength increase after whole-body vibration compared with resistance training. *Med Sci Sports Exerc* 2003;**35**:1033–1041.

47. Roelants M, Delecluse C, Verschueren SM. Whole-body-vibration training increases knee-extension strength and speed of movement in older women. *J Am Geriatr Soc* 2004;**52**:901–908.

48. Phillips SM. Nutritional Supplements in Support of Resistance Exercise to Counter Age-Related Sarcopenia. *Adv Nutr* 2015;**6**:452–460.

49. Rodacki CL, Rodacki AL, Pereira G, Naliwaiko K, Coelho I, Pequito D *et al.* Fish-oil supplementation enhances the effects of strength training in elderly women. *Am J Clin Nutr* 2012;**95**:428–436.

50. Cermak NM, Res PT, de Groot LC, Saris WH, van Loon LJ. Protein supplementation augments the adaptive response of skeletal muscle to resistance-type exercise training: a meta-analysis. *Am J Clin Nutr* 2012;**96**:1454–1464.

51. Vellers HL, Kleeberger SR, Lightfoot JT. Inter-individual variation in adaptations to endurance and resistance exercise training: genetic approaches towards understanding a complex phenotype. *Mamm Genome* 2018;**29**:48–62.

52. Ahtiainen JP, Walker S, Peltonen H, Holviala J, Sillanpää E, Karavirta L *et al.* Heterogeneity in resistance training-induced muscle strength and mass responses in men and women of different ages. *Age (Omaha)* 2016;**38**:1–13.

53. Hubal MJ, Gordish-Dressman H, Thompson PD, Price TB, Hoffman EP, Angelopoulos TJ *et al.* Variability in muscle size and strength gain after unilateral resistance training. *Med Sci Sports Exerc* 2005;**37**:964–972.

54. Thomis MA, Beunen GP, Maes HH, Blimkie CJ, Van Leemputte M, Claessens AL *et al.* Strength training: importance of genetic factors. *Med Sci Sports Exerc* 1998;**30**:724–31.

55. Thompson PD, Moyna N, Seip R, Price T, Clarkson P, Angelopoulos T *et al.* Functional Polymorphisms Associated with Human Muscle Size and Strength. *Med Sci Sport Exerc* 2004;**36**:1132–1139.

56. Pescatello LS, Devaney JM, Hubal MJ, Thompson PD, Hoffman EP. Highlights from the Functional Single Nucleotide Polymorphisms Associated with Human Muscle Size and Strength or FAMuSS Study. *Biomed Res Int* 2013;**2013**:1–11.

57. Mithal A, Bonjour JP, Boonen S, Burckhardt P, Degens H, El Hajj Fuleihan G *et al.* Impact of nutrition on muscle mass, strength, and performance in older adults. *Osteoporos Int* 2013;**24**:1555–1566.

58. Arden NK, Spector TD. Genetic influences on muscle strength, lean body mass, and bone mineral density: A twin study. *J Bone Miner Res* 1997;**12**:2076–2081.

59. Carmelli D, Reed T. Stability and change in genetic and environmental influences on hand-grip strength in older male twins. *J Appl Physiol* 2000;**89**:1879–1883.

60. Frederiksen H, Gaist D, Christian Petersen H, Hjelmborg J, McGue M, Vaupel JW *et al.* Hand grip strength: A phenotype suitable for identifying genetic variants affecting mid- and late-life physical functioning. *Genet Epidemiol* 2002;**23**:110–122.

61. Tiainen K, Sipilä S, Kauppinen M, Kaprio J, Rantanen T. Genetic and environmental effects on isometric muscle strength and leg extensor power

followed up for three years among older female twins. *J Appl Physiol* 2009;**106**:1604–1610.

62. Zempo H, Miyamoto-Mikami E, Kikuchi N, Fuku N, Miyachi M, Murakami H. Heritability estimates of muscle strength-related phenotypes: A systematic review and meta-analysis. *Scand J Med Sci Sports* 2017;**27**:1537–1546.

63. Caló MC, Vona G. Gene polymorphisms and elite athletic performance. *J Anthropol Sci* 2008;**86**:113–31.

64. Bray MS, Hagberg JM, Pérusse L, Rankinen T, Roth SM, Wolfarth B *et al. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update.* 2009.

65. Puthucheary Z, Skipworth JR a, Rawal J, Loosemore M, Van Someren K, Montgomery HE. Genetic influences in sport and physical performance. *Sport Med* 2011;**41**:845–59.

66. Garatachea N, Lucía A. Genes and the ageing muscle: A review on genetic association studies. *Age (Omaha)* 2013;**35**:207–233.

67. Cho J, Lee I, Kang H. ACTN3 Gene and Susceptibility to Sarcopenia and Osteoporotic Status in Older Korean Adults. *Biomed Res Int* 2017;**2017**:1–8.

68. Williams AG, Day SH, Folland JP, Gohlke P, Dhamrait S, Montgomery HE. Circulating angiotensin converting enzyme activity is correlated with muscle strength. *Med Sci Sports Exerc* 2005;**37**:944–948.

69. Charbonneau DE, Hanson ED, Ludlow AT, Delmonico MJ, Hurley BF, Roth SM. ACE genotype and the muscle hypertrophic and strength responses to strength training. *Med Sci Sports Exerc* 2008;**40**:677–683.

70. Arking DE, Fallin DM, Fried LP, Li T, Beamer BA, Xue QL *et al.* Variation in the Ciliary Neurotrophic Factor Gene and Muscle Strength in Older Caucasian Women. *J Am Geriatr Soc* 2006;**54**:823–826.

71. Heffernan SM, Stebbings GK, Kilduff LP, Erskine RM, Day SH, Morse CI *et al.* Fat mass and obesity associated (FTO) gene influences skeletal muscle phenotypes in non-resistance trained males and elite rugby playing position. *BMC Genet* 2017;**18**:1–9.

72. Drozdovska SB, Dosenko VE, Ahmetov II, Ilyin VN. The association of gene

polymorphisms with athlete status in Ukrainians. *Biol Sport* 2013;**30**:163–167.

73. Gabbasov RT, Arkhipova AA, Borisova A V., Hakimullina AM, Kuznetsova A V., Williams AG *et al.* The HIF1A Gene Pro582Ser Polymorphism in Russian Strength Athletes. *J Strength Cond Res* 2013;**27**:2055–2058.

74. Cięszczyk P, Kalinski M, Ostanek M, Jascaniene N, Krupecki K, Ficek K *et al.* Variation in the HIF1A Gene in Elite Rowers. *J Strength Cond Res* 2012;**26**:3270–3274.

75. Seibert MJ, Xue QL, Fried LP, Walston JD. Polymorphic variation in the human myostatin (GDF-8) gene and association with strength measures in the women's health and aging study II cohort. *J Am Geriatr Soc* 2001;**49**:1093–1096.

76. González-Freire M, Rodríguez-Romo G, Santiago C, Bustamante-Ara N, Yvert T, Gómez-Gallego F *et al.* The K153R variant in the myostatin gene and sarcopenia at the end of the human lifespan. *Age (Omaha)* 2010;**32**:405–409.

77. Santiago C, Ruiz JR, Rodríguez-Romo G, Fiuza-Luces C, Yvert T, Gonzalez-Freire M *et al.* The K153R Polymorphism in the Myostatin Gene and Muscle Power Phenotypes in Young, Non-Athletic Men. *PLoS One* 2011;**6**:1–5.

78. Roth SM, Zmuda JM, Cauley JA, Shea PR, Ferrell RE. Vitamin D Receptor Genotype Is Associated With Fat-Free Mass and Sarcopenia in Elderly Men. *Journals Gerontol Ser A Biol Sci Med Sci* 2004;**59**:B10–B15.

79. Windelinckx A, De Mars G, Beunen G, Aerssens J, Delecluse C, Lefevre J *et al.* Polymorphisms in the vitamin D receptor gene are associated with muscle strength in men and women. *Osteoporos Int* 2007;**18**:1235–1242.

80. Hopkinson NS, Li KW, Kehoe A, Humphries SE, Roughton M, Moxham J *et al.* Vitamin D receptor genotypes influence quadriceps strength in chronic obstructive pulmonary disease. *Am J Clin Nutr* 2008;**87**:385–390.

81. McLeod M, Breen L, Hamilton DL, Philp A. Live strong and prosper: the importance of skeletal muscle strength for healthy ageing. *Biogerontology* 2016;**17**:1–14.

82. Ni Lochlainn M, Bowyer R, Steves C. Dietary Protein and Muscle in Aging People: The Potential Role of the Gut Microbiome. *Nutrients* 2018;**10**:929.

83. Wu Ct, Morris JR. Genes, genetics, and epigenetics: a correspondence. *Science* 2001;**293**:1103–5.

84. Brown WM. Exercise-associated DNA methylation change in skeletal muscle and the importance of imprinted genes: A bioinformatics meta-analysis. *Br J Sports Med* 2015;**49**:1568–1578.

85. Chen Z, Li S, Subramaniam S, Shyy JY-J, Chien S. Epigenetic Regulation: A New Frontier for Biomedical Engineers. *Annu Rev Biomed Eng* 2017;**19**:195–219.

86. Stephens KE, Miaskowski CA, Levine JD, Pullinger CR, Aouizerat BE. Epigenetic Regulation and Measurement of Epigenetic Changes. *Biol Res Nurs* 2013;**15**:373–381.

87. Zhang N. Epigenetic modulation of DNA methylation by nutrition and its mechanisms in animals. *Anim Nutr* 2015;**1**:144–151.

88. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* 1982;**10**:2709–21.

89. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci* 2006;**103**:1412–1417.

90. Zeng J, Nagrajan HK, Yi S V. Fundamental diversity of human CpG islands at multiple biological levels. *Epigenetics* 2014;**9**:483–491.

91. Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. *Aging Cell* 2015;**14**:924–932.

92. Zuo Z, Roy B, Chang YK, Granas D, Stormo GD. Measuring quantitative effects of methylation on transcription factor–DNA binding affinity. *Sci Adv* 2017;**3**:eaao1799.

93. Lorincz MC, Dickerson DR, Schmitt M, Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol* 2004;**11**:1068–1075.

94. Jeziorska DM, Murray RJS, De Gobbi M, Gaentzsch R, Garrick D, Ayyub H *et al.* DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proc Natl Acad Sci* 2017;**114**:E7526–E7535.

95. Lokk K, Modhukur V, Rajashekar B, Märtens K, Mägi R, Kolde R *et al.* DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol* 2014;**15**:3248.

96. Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. *Epigenetics and Chromatin* 2018;**11**:1–11.

97. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013;**14**:R115.

98. Jung M, Pfeifer GP. Aging and DNA methylation. *BMC Biol* 2015;**13**:1–8.

99. Unnikrishnan A, Hadad N, Masser DR, Jackson J, Freeman WM, Richardson A. Revisiting the genomic hypomethylation hypothesis of aging. *Ann N Y Acad Sci* 2018;**1418**:69–79.

100. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S *et al.* Epigenetic Predictor of Age. *PLoS One* 2011;**6**:e14821.

101. Zykovich A, Hubbard A, Flynn JM, Tarnopolsky M, Fraga MF, Kerksick C *et al.* Genome-wide DNA methylation changes with age in disease-free human skeletal muscle. *Aging Cell* 2014;**13**:360–366.

102. Seaborne RA, Strauss J, Cocks M, Shepherd S, O'Brien TD, van Someren KA *et al.* Human Skeletal Muscle Possesses an Epigenetic Memory of Hypertrophy. *Sci Rep* 2018;**8**:1898.

103. Sharples AP, Stewart CE, Seaborne RA. Does skeletal muscle have an 'epi'-memory? The role of epigenetics in nutritional programming, metabolic disease, aging and exercise. *Aging Cell* 2016;**15**:603–616.

104. Turner DC, Seaborne RA, Sharples AP. Comparative Transcriptome and Methylome Analysis in Human Skeletal Muscle Anabolism, Hypertrophy and Epigenetic Memory. *Sci Rep* 2019;**9**:1–12.

105. Sailani MR, Halling JF, Møller HD, Lee H, Plomgaard P, Pilegaard H *et al.*

Lifelong physical activity is associated with promoter hypomethylation of genes involved in metabolism, myogenesis, contractile properties and oxidative stress resistance in aged human skeletal muscle. *Sci Rep* 2019;**9**:1–11.

106. Posthuma D, Beem AL, de Geus EJC, van Baal GCM, von Hjelmborg JB, Iachine I *et al.* Theory and Practice in Quantitative Genetics. *Twin Res* 2003;**6**:361–376.

107. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974;**19**:716–723.

108. Thomis MA, Van Leemputte M, Maes HH, Blimkie CJR, Claessens AL, Marchal G *et al.* Multivariate genetic analysis of maximal isometric muscle force at different elbow angles. *J Appl Physiol* 1997;**82**:959–967.

109. Heckerman D, Traynor BJ, Picca A, Calvani R, Marzetti E, Hernandez D *et al.* Genetic variants associated with physical performance and anthropometry in old age: a genome-wide association study in the ilSIRENTE cohort. *Sci Rep* 2017;**7**:15879.

110. Tikkanen E, Gustafsson S, Amar D, Shcherbina A, Waggott D, Ashley EA *et al.* Biological Insights Into Muscular Strength: Genetic Findings in the UK Biobank. *Sci Rep* 2018;**8**:6451.

111. Liu XG, Tan LJ, Lei SF, Liu YJ, Shen H, Wang L *et al.* Genome-wide Association and Replication Studies Identified TRHR as an Important Gene for Lean Body Mass. *Am J Hum Genet* 2009;**84**:418–423.

112. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30.

113. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;**32**:D258-61.

114. Welle S, Bhatt K, Pinkert CA, Tawil R, Thornton CA. Muscle growth after postdevelopmental myostatin gene knockout. *Am J Physiol Metab* 2007;**292**:E985–E991.

115. Broos S, Malisoux L, Theisen D, van Thienen R, Ramaekers M, Jamart C *et al.* Evidence for ACTN3 as a Speed Gene in Isolated Human Muscle Fibers.

PLoS One 2016;**11**:e0150594.

116. Choi SW, Mak TSH, O'Reilly PF. A guide to performing Polygenic Risk Score analyses. *bioRxiv* 2018;**2**:416545.

117. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* 2019;**10**:3328.

118. Williams AG, Folland JP. Similarity of polygenic profiles limits the potential for elite human physical performance. *J Physiol* 2008;**586**:113–121.

119. Ruiz JR, Gómez-Gallego F, Santiago C, González-Freire M, Verde Z, Foster C *et al.* Is there an optimum endurance polygenic profile? *J Physiol* 2009;**587**:1527–1534.

120. Santiago C, Ruiz JR, Muniesa CA, González-Freire M, Gómez-Gallego F, Lucia A. Does the polygenic profile determine the potential for becoming a world-class athlete? Insights from the sport of rowing. *Scand J Med Sci Sport* 2010;**20**:e188–e194.

121. Thomaes T, Thomis M, Onkelinx S, Fagard R, Matthijs G, Buys R *et al.* A genetic predisposition score for muscular endophenotypes predicts the increase in aerobic power after training: The CAREGENE study. *BMC Genet* 2011;**12**:84–93.

122. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ *et al.* Genomic predictors of the maximal O2 uptake response to standardized exercise training programs. *J Appl Physiol* 2011;**110**:1160–1170.

123. Thomaes T, Thomis M, Onkelinx S, Goetschalckx K, Fagard R, Lambrechts D *et al.* Genetic predisposition scores associate with muscular strength, size, and trainability. *Med Sci Sports Exerc* 2013;**45**:1451–1459.

124. Charlier R, Caspers M, Knaeps S, Mertens E, Lambrechts D, Lefevre J *et al.* Limited potential of genetic predisposition scores to predict muscle mass and strength performance in Flemish Caucasians between 19 and 73 years of age. *Physiol Genomics* 2017;**49**:160–166.

125. Wei JH, Haddad A, Wu KJ, Zhao HW, Kapur P, Zhang ZL *et al.* A CpG-methylation-based assay to predict survival in clear cell renal cell carcinoma. *Nat Commun* 2015;**6**:1–11.

126. Ahmad AS, Vasiljević N, Carter P, Berney DM, Møller H, Foster CS *et al.* A novel DNA methylation score accurately predicts death from prostate cancer in men with low to intermediate clinical risk factors. *Oncotarget* 2016;**7**.

127. Zhao S, Leonardson A, Geybels MS, McDaniel AS, Yu M, Kolb S *et al.* A five-CpG DNA methylation score to predict metastatic-lethal outcomes in men treated with radical prostatectomy for localized prostate cancer. *Prostate* 2018;**78**:1084–1091.

128. Hamilton OKL, Zhang Q, McRae AF, Walker RM, Morris SW, Redmond P *et al.* An epigenetic score for BMI based on DNA methylation correlates with poor physical health and major disease in the Lothian Birth Cohort. *Int J Obes* 2019 doi:10.1038/s41366-018-0262-3.

129. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;**11**:587.

130. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A *et al.* Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *Am J Hum Genet* 2015;**97**:75–85.

131. Dogan M V., Grumbach IM, Michaelson JJ, Philibert RA. Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study. *PLoS One* 2018;**13**:1–18.

132. Reese SE, Zhao S, Wu MC, Joubert BR, Parr CL, Håberg SE *et al.* DNA Methylation Score as a Biomarker in Newborns for Sustained Maternal Smoking during Pregnancy. *Environ Health Perspect* 2017;**125**:760–766.

133. Cho S, Kim K, Kim YJ, Lee JK, Cho YS, Lee JY *et al.* Joint Identification of Multiple Genetic Variants via Elastic-Net Variable Selection in a Genome-Wide Association Analysis. *Ann Hum Genet* 2010;**74**:416–428.

134. Sakuma K, Yamaguchi A. Sarcopenia and Age-Related Endocrine Function. *Int J Endocrinol* 2012;**2012**:1–10.

135. McKee A, Morley JE, Matsumoto AM, Vinik A. SARCOPENIA: AN ENDOCRINE DISORDER? *Endocr Pract* 2017;**23**:1143–1152.

136. Kwak JY, Hwang H, Kim SK, Choi JY, Lee SM, Bang H *et al.* Prediction of sarcopenia using a combination of multiple serum biomarkers. *Sci Rep*

2018;**8**:1–7.

137. Bogaerts A, Delecluse C, Claessens AL, Coudyzer W, Boonen S, Verschueren SMP. Impact of whole-body vibration training versus fitness training on muscle strength and muscle mass in older men: A 1-year randomized controlled trial. *Journals Gerontol Ser A Biol Sci Med Sci* 2007;**62**:630–635.

138. Bogaerts A, Delecluse C, Claessens AL, Troosters T, Boonen S, Verschueren SMP. Effects of whole body vibration training on cardiorespiratory fitness and muscle strength in older individuals (a 1-year randomised controlled trial). *Age Ageing* 2009;**38**:448–454.

139. Nica AC, Dermitzakis ET. Expression quantitative trait loci: Present and future. *Philos Trans R Soc B Biol Sci* 2013;**368**.

140. Janssen I, Heymsfield SB, Baumgartner RN, Ross R. Estimation of skeletal muscle mass by bioelectrical impedance analysis. *J Appl Physiol* 2000;**89**:465–471.

141. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;**30**:1363–1369.

142. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT *et al.* FlowSorted.Blood.EPIC. Bioconductor. 2018 doi:10.18129/B9.bioc.FlowSorted.Blood.EPIC.

143. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* 2016;**17**:208.

144. Ogasawara R, Thiebaud RS, Loenneke JP, Loftin M, Abe T. Time course for arm and chest muscle thickness changes following bench press training. *Interv Med Appl Sci* 2012;**4**:217–220.

145. Reeves ND, Maganaris CN, Narici M V. Ultrasonographic assessment of human skeletal muscle size. *Eur J Appl Physiol* 2004;**91**:116–118.

146. Visser M, Kritchevsky S, Goodpaster B, Newman AB, Nevitt MC, Stamm E *et al.* Leg muscle mass and composition in relation to lower extremity

performance in men and women aged 70 to 79: The health, aging and body composition study. *J Am Geriatr Soc* 2002;**50**:897–904.

147. Short KR, Bigelow ML, Kahl J, Singh R, Coenen-Schimke J, Raghavakaimal S *et al.* Decline in skeletal muscle mitochondrial function with aging in humans. *Proc Natl Acad Sci* 2005;**102**:5618–5623.

148. Kanegusuku H, Queiroz ACC, Silva VJD, De Mello MT, Ugrinowitsch C, Forjaz CLM. High-intensity progressive resistance training increases strength with no change in cardiovascular function and autonomic neural regulation in older adults. *J Aging Phys Act* 2015;**23**:339–345.

149. Liberman K, Forti LN, Beyer I, Bautmans I. The effects of exercise on muscle strength, body composition, physical functioning and the inflammatory profile of older adults: a systematic review. *Curr Opin Clin Nutr Metab Care* 2017;**20**:30–53.

150. Cadore E, Pinto R, Bottaro M, Mikel I. Strength and Endurance Training Prescription in Healthy and Frail Elderly. *Aging Dis* 2014;**5**:183.

151. Machado A, García-López D, González-Gallego J, Garatachea N. Whole-body vibration training increases muscle strength and mass in older women: A randomized-controlled trial. *Scand J Med Sci Sport* 2010;**20**:200–207.

152. Kennis E, Verschueren SM, Bogaerts A, Coudyzer W, Boonen S, Delecluse C. Effects of fitness and vibration training on muscle quality: A 1-year postintervention follow-up in older men. *Arch Phys Med Rehabil* 2013;**94**:910–918.

153. Bouchard C, Dionne FT, Simoneau JA, Boulay MR. Genetics of aerobic and anaerobic performances. *Exerc Sport Sci Rev* 1992;**20**:27–58.

154. Montgomery HE, Marshall R, Hemingway H, Myerson S, Clarkson P, Dollery C *et al.* Human gene for physical performance. *Nature* 1998;**393**:221–222.

155. Papadimitriou ID, Lucia A, Pitsiladis YP, Pushkarev VP, Dyatlov DA, Orekhov EF *et al.* ACTN3 R577X and ACE I/D gene variants influence performance in elite sprinters: A multi-cohort study. *BMC Genomics* 2016;**17**:1–8.

156. Petr M, Šťastný P, Pecha O, Šteffl M, Šeda O, Kohlíková E. PPARA intron polymorphism associated with power performance in 30-s anaerobic wingate

test. *PLoS One* 2014;**9**:1–5.

157. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ *et al.* Genomic predictors of the maximal O2 uptake response to standardized exercise training programs. *J Appl Physiol* 2011;**110**:1160–1170.

158. American College of Sports Medicine. *ACSM's Guidelines for exercise testing and prescription*. 6th ed. Philadelphia (Pa.) : Lippincott Williams and Wilkins; 2000.

159. Windelinckx A, De Mars G, Huygens W, Peeters MW, Vincent B, Wijmenga C *et al.* Comprehensive fine mapping of chr12q12-14 and follow-up replication identify activin receptor 1B (ACVR1B) as a muscle strength gene. *Eur J Hum Genet* 2011;**19**:208–215.

160. Voisin S, Guilherme JPFL, Yan X, Pushkarev VP, Cieszczyk P, Massidda M *et al.* ACVR1B rs2854464 is associated with sprint/power athletic status in a large cohort of Europeans but not Brazilians. *PLoS One* 2016;**11**:1–11.

161. Miyamoto-Mikami E, Murakami H, Tsuchie H, Takahashi H, Ohiwa N, Miyachi M *et al.* Lack of association between genotype score and sprint/power performance in the Japanese population. *J Sci Med Sport* 2017;**20**:98–103.

162. Foulstone EJ, Savage PB, Crown AL, Holly JMP, Stewart CEH. Role of insulin-like growth factor binding protein-3 (IGFBP-3) in the differentiation of primary human adult skeletal myoblasts. *J Cell Physiol* 2003;**195**:70–79.

163. Cheng I, Henderson KD, Haiman CA, Kolonel LN, Henderson BE, Freedman ML *et al.* Genetic determinants of circulating insulin-like growth factor (IGF)-I, IGF binding protein (BP)-1, and IGFBP-3 levels in a multiethnic population. *J Clin Endocrinol Metab* 2007;**92**:3660–3666.

164. Joulia-Ekaza D, Cabello G. The myostatin gene: physiology and pharmacological relevance. *Curr Opin Pharmacol* 2007;**7**:310–315.

165. Lee S-J, McPherron AC. Regulation of myostatin activity and muscle growth. *Proc Natl Acad Sci* 2001;**98**:9306–9311.

166. Walsh S, Metter EJ, Ferrucci L, Roth SM. Activin-type II receptor B (ACVR2B) and follistatin haplotype associations with muscle mass and strength in humans. *J Appl Physiol* 2007;**102**:2142–2148.

167. Stebbings GK, Williams AG, Herbert AJ, Lockey SJ, Heffernan SM, Erskine RM *et al.* TTN genotype is associated with fascicle length and marathon running performance. *Scand J Med Sci Sport* 2017;1–7.

168. Wang Y, DeLuca HF. Is the vitamin D receptor found in muscle? *Endocrinology* 2011;**152**:354–363.

169. Bozsodi A, Boja S, Szilagyi A, Somhegyi A, Varga PP, Lazary A. Muscle strength is associated with vitamin D receptor gene variants. *J Orthop Res* 2016;**34**:2031–2037.

170. Amthor H, Macharia R, Navarrete R, Schuelke M, Brown SC, Otto A *et al.* Lack of myostatin results in excessive muscle growth but impaired force generation. *Proc Natl Acad Sci U S A* 2007;**104**:1835–40.

171. Schiaffino S, Mammucari C. Regulation of skeletal muscle growth by the IGF1-Akt/PKB pathway: Insights from genetic models. *Skelet Muscle* 2011;**1**:4.

172. Rommel C, Bodine S, Clarke B. Mediation of IGF-1-induced skeletal myotube hypertrophy by PI (3) K/Akt/mTOR and PI (3) K/Akt/GSK3 pathways. *Nat Cell Biol* 2001;**3**:1009.

173. Akasaki Y, Ouchi N, Izumiya Y, Bernardo BL, Lebrasseur NK, Walsh K. Glycolytic fast-twitch muscle fiber restoration counters adverse age-related changes in body composition and metabolism. *Aging Cell* 2014;**13**:80–91.

174. Hubal MJ, Devaney JM, Hoffman EP, Zambraski EJ, Gordish-Dressman H, Kearns AK *et al.* CCL2 and CCR2 polymorphisms are associated with markers of exercise-induced skeletal muscle damage. *J Appl Physiol* 2010;**108**:1651–1658.

175. Harmon BT, Orkunoglu-Suer EF, Adham K, Larkin JS, Gordish-Dressman H, Clarkson PM *et al.* CCL2 and CCR2 variants are associated with skeletal muscle strength and change in strength with resistance training. *J Appl Physiol* 2010;**109**:1779–85.

176. Van Rossum EFC, Voorhoeve PG, te Velde SJ, Koper JW, Delemarre-van de Waal HA, Kemper HCG *et al.* The ER22/23EK polymorphism in the glucocorticoid receptor gene is associated with a beneficial body composition and muscle strength in young adults. *J Clin Endocrinol Metab*

2004;**89**:4004–4009.

177. Ash GI, Kostek MA, Lee H, Angelopoulos TJ, Clarkson PM, Gordon PM *et al.* Glucocorticoid receptor (NR3C1) variants associate with the muscle strength and size response to resistance training. *PLoS One* 2016;**11**:1–11.

178. Huang J, Hsu Y-H, Mo C, Abreu E, Kiel DP, Bonewald LF *et al.* METTL21C Is a Potential Pleiotropic Gene for Osteoporosis and Sarcopenia Acting Through the Modulation of the NF-κB Signaling Pathway. *J Bone Miner Res* 2014;**29**:1531–1540.

179. Hangelbroek RWJ, Fazelzadeh P, Tieland M, Boekschoten M V., Hooiveld GJEJ, van Duynhoven JPM *et al.* Expression of protocadherin gamma in skeletal muscle tissue is associated with age and muscle weakness. *J Cachexia Sarcopenia Muscle* 2016;**7**:604–614.

180. Pistilli EE, Bogdanovich S, Garton F, Yang N, Gulbin JP, Conner JD *et al.* Loss of IL-15 receptor alpha alters the endurance, fatigability, and metabolic characteristics of mouse fast skeletal muscles. *J Clin Invest* 2011;**121**:3120–3132.

181. Pistilli EE, Devaney JM, Gordish-Dressman H, Bradbury MK, Seip RL, Thompson PD *et al.* Interleukin-15 and interleukin-15Rα SNPs and associations with muscle, bone, and predictors of the metabolic syndrome. *Cytokine* 2008;**43**:45–53.

182. Kyriakides T. SPP1 genotype is a determinant of disease severity in Duchenne muscular dystrophy: Predicting the severity of duchenne muscular dystrophy: Implications for treatment. *Neurology* 2011;**77**:1858.

183. Voisin S, Eynon N, Yan X, Bishop DJ. Exercise training and DNA methylation in humans. *Acta Physiol* 2015;**213**:39–59.

184. Terruzzi I, Senesi P, Montesano A, Torre A La, Alberti G, Benedini S *et al.* Genetic polymorphisms of the enzymes involved in DNA methylation and synthesis in elite athletes. *Physiol Genomics* 2011;**43**:965–973.

185. Zarebska A, Ahmetov II, Sawczyn S, Weiner AS, Kaczmarczyk M, Ficek K *et al.* Association of the MTHFR 1298A>C (rs1801131) polymorphism with speed and strength sports in Russian and Polish athletes. *J Sports Sci* 2014;**32**:375–382.

186. Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem* 2004;**279**:27816–27823.

187. Van Roie E, Walker S, Van Driessche S, Baggen R, Coudyzer W, Bautmans I *et al.* Training load does not affect detraining's effect on muscle volume, muscle strength and functional capacity among older adults. *Exp Gerontol* 2017;**98**:30–37.

188. Massidda M, Scorcu M, Calò CM. New genetic model for predicting phenotype traits in sports. *Int J Sports Physiol Perform* 2014;**9**:554–560.

189. Abraham G, Kowalczyk A, Zobel J, Inouye M. Performance and robustness of penalized and unpenalized methods for genetic prediction of complex human disease. *Genet Epidemiol* 2013;**37**:184–195.

190. Ghosh S, Vivar JC, Sarzynski MA, Sung YJ, Timmons JA, Bouchard C *et al.* Integrative pathway analysis of a genome-wide association study of VO2max response to exercise training. *J Appl Physiol* 2013;**115**:1343–1359.

191. Willems SM, Wright DJ, Day FR, Trajanoska K, Joshi PK, Morris JA *et al.* Large-scale GWAS identifies multiple loci for hand grip strength providing biological insights into muscular fitness. *Nat Commun* 2017;**8**.

192. Timmons JA, Knudsen S, Rankinen T, Koch LG, Sarzynski M, Jensen T *et al.* Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. *J Appl Physiol* 2010;**108**:1487–1496.

193. Bird A. Perceptions of epigenetics. *Nature* 2007;**447**:396–398.

194. Hanson MA, Gluckman PD. Early developmental conditioning of later health and disease: Physiology or pathophysiology? *Physiol Rev* 2014;**94**:1027–1076.

195. Barrès R, Yan J, Egan B, Treebak JT, Rasmussen M, Fritz T *et al.* Acute exercise remodels promoter methylation in human skeletal muscle. *Cell Metab* 2012;**15**:405–411.

196. Pitsiladis YP, Tanaka M, Eynon N, Bouchard C, North KN, Williams AG *et al.* Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance. *Physiol Genomics*

2016;**48**:183–190.

197. Yan X, Eynon N, Papadimitriou ID, Kuang J, Munson F, Tirosh O *et al.* The gene SMART study: Method, study design, and preliminary findings. *BMC Genomics* 2017;**18**.

198. Sayer AA, Syddall H, Martin H, Patel H, Baylis D. The Developmental Origins of Ageing. *J Nutr Heal aging* 2008;**12**:427–432.

199. Lopez P, Pinto RS, Radaelli R, Rech A, Grazioli R, Izquierdo M *et al.* Benefits of resistance training in physically frail elderly: a systematic review. *Aging Clin Exp Res* 2018;**30**:889–899.

200. Guizelini PC, de Aguiar RA, Denadai BS, Caputo F, Greco CC. Effect of resistance training on muscle strength and rate of force development in healthy older adults: A systematic review and meta-analysis. *Exp Gerontol* 2018;**102**:51–58.

201. Sitjà-Rabert M, Rigau D, Fort Vanmeerghaeghe A, Romero-Rodríguez D, Bonastre Subirana M, Bonfill X. Efficacy of whole body vibration exercise in older people: A systematic review. *Disabil Rehabil* 2012;**34**:883–893.

202. Karavirta L, Häkkinen A, Sillanpää E, García-López D, Kauhanen A, Haapasaari A *et al.* Effects of combined endurance and strength training on muscle strength, power and hypertrophy in 40-67-year-old men. *Scand J Med Sci Sports* 2011;**21**:402–11.

203. Fatouros IG, Kambas A, Katrabasas I, Nikolaidis K, Chatzinikolaou A, Leontsini D *et al.* Strength training and detraining effects on muscular strength, anaerobic power, and mobility of inactive older men are intensity dependent. *Br J Sports Med* 2005;**39**:776–780.

204. Correa CS, Baroni BM, Radaelli R, Lanferdini FJ, Dos Santos Cunha G, Reischak-Oliveira Á *et al.* Effects of strength training and detraining on knee extensor strength, muscle volume and muscle quality in elderly women. *Age (Omaha)* 2013;**35**:1899–1904.

205. Taaffe DR, Henwood TR, Nalls MA, Walker DG, Lang TF, Harris TB. Alterations in muscle attenuation following detraining and retraining in resistance-trained older adults. *Gerontology* 2009;**55**:217–223.

206. Taaffe DR, Marcus R. Dynamic muscle strength alterations to detraining and

retraining in elderly men. *Clin Physiol* 1997;**17**:311–324.

207. Blazevich AJ. Effects of Physical Training and Detraining, Immobilisation, Growth and Aging on Human Fascile Geometry. *Sports Med* 2006;**36**:1003–1017.

208. Andersen LL, Andersen JL, Magnusson SP, Aagaard P. Neuromuscular adaptations to detraining following resistance training in previously untrained subjects. *Eur J Appl Physiol* 2005;**93**:511–518.

209. Lovell DI, Cuneo R, Wallace J, McLellan C. The hormonal response of older men to sub-maximum aerobic exercise: The effect of training and detraining. *Steroids* 2012;**77**:413–418.

210. Bray MS, Hagberg JM, Pérusse L, Rankinen T, Roth SM, Wolfarth B *et al.* The human gene map for performance and health-related fitness phenotypes: The 2006-2007 update. *Med Sci Sports Exerc* 2009;**41**:34–72.

211. Delmonico MJ, Kostek MC, Doldo NA, Hand BD, Walsh S, Conway JM *et al.* Alpha-Actinin-3 (ACTN3) R577X Polymorphism Influences Knee Extensor Peak Power Response to Strength Training in Older Men and Women. *J Gerontol A Biol Sci Med Sci* 2007;**62**:206–212.

212. He L, Van Roie E, Bogaerts A, Morse CI, Delecluse C, Verschueren S *et al.* Genetic predisposition score predicts the increases of knee strength and muscle mass after one-year exercise in healthy elderly. *Exp Gerontol* 2018;**111**:17–26.

213. Illumina. GoldenGate® Assay Workflow. 2006https://www.illumina.com/Documents/products/workflows/workflow_goldengate_assay.pdf.

214. Liu K, Zhao R, Shen M, Ye J, Li X, Huang Y *et al.* Role of genetic mutations in folate-related enzyme genes on Male Infertility. *Sci Rep* 2015;**5**:15548.

215. Ahmetov II, Mozhayskaya IA, Flavell DM, Astratenkova I V., Komkova AI, Lyubaeva E V. *et al.* PPARα gene variation and physical performance in Russian athletes. *Eur J Appl Physiol* 2006;**97**:103–108.

216. Walsh S, Kelsey BK, Angelopoulos TJ, Clarkson PM, Gordon PM, Moyna NM *et al.* CNTF 1357 G → A polymorphism and the muscle strength response to resistance training. *J Appl Physiol* 2009;**107**:1235–1240.

217. De Mars G, Windelinckx A, Beunen G, Delecluse C, Lefevre J, Thomis M a I. Polymorphisms in the CNTF and CNTF receptor genes are associated with muscle strength in men and women. *J Appl Physiol* 2007;**102**:1824–1831.

218. Delmonico MJ, Zmuda JM, Taylor BC, Cauley JA, Harris TB, Manini TM *et al.* Association of the ACTN3 genotype and physical functioning with age in older adults. *Journals Gerontol - Ser A Biol Sci Med Sci* 2008;**63**:1227–1234.

219. Schrager M a, Roth SM, Ferrell RE, Metter EJ, Russek-Cohen E, Lynch N a *et al.* Insulin-like growth factor-2 genotype, fat-free mass, and muscle performance across the adult life span. *J Appl Physiol* 2004;**97**:2176–2183.

220. Massidda M, Scorcu M, Calò CM. New genetic model for predicting phenotype traits in sports. *Int J Sports Physiol Perform* 2014;**9**:554–560.

221. Dupont C, Armant D, Brenner C. Epigenetics: Definition, Mechanisms and Clinical Perspective. *Semin Reprod Med* 2009;**27**:351–357.

222. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *J Mol Biol* 1987;**196**:261–82.

223. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev* 2011;**25**:1010–1022.

224. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 2017;**543**:72–77.

225. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* 2016;**8**:389–399.

226. Rao X, Evans J, Chae H, Pilrose J, Kim S, Yan P *et al.* CpG island shore methylation regulates caveolin-1 expression in breast cancer. *Oncogene* 2013;**32**:4519–4528.

227. Martin EM, Fry RC. Environmental Influences on the Epigenome: Exposure-Associated DNA Methylation in Human Populations. *Annu Rev Public Health* 2018;**39**:309–333.

228. Nitert MD, Dayeh T, Volkov P, Elgzyri T, Hall E, Nilsson E *et al.* Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. *Diabetes* 2012;**61**:3322–3332.

229. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci* 2012;**109**:10522–10527.

230. Bell JT, Tsai P, Yang T, Pidsley R, Nisbet J, Glass D *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012;**8**:e1002629.

231. Bormann F, Rodríguez-Paredes M, Hagemann S, Manchanda H, Kristof B, Gutekunst J *et al.* Reduced DNA methylation patterning and transcriptional connectivity define human skin aging. *Aging Cell* 2016;**15**:563–571.

232. Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der brug M, Chong S *et al.* Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet* 2011;**20**:1164–1172.

233. Sharples AP, Polydorou I, Hughes DC, Owens DJ, Hughes TM, Stewart CE. Skeletal muscle cells possess a 'memory' of acute early life TNF-α exposure: role of epigenetic adaptation. *Biogerontology* 2016;**17**:603–617.

234. Mahoney SE, Yao Z, Keyes CC, Tapscott SJ, Diede SJ. Genome-wide DNA methylation studies suggest distinct DNA methylation patterns in pediatric embryonal and alveolar rhabdomyosarcomas. *Epigenetics* 2012;**7**:400–8.

235. Wang M, Xie H, Shrestha S, Sredni S, Morgan GA, Pachman LM. Methylation alterations of WT1 and homeobox genes in inflamed muscle biopsy samples from patients with untreated juvenile dermatomyositis suggest self-renewal capacity. *Arthritis Rheum* 2012;**64**:3478–3485.

236. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 1995;**57**:289–300.

237. Tan LJ, Liu SL, Lei SF, Papasian CJ, Deng HW. Molecular genetic studies of gene identification for sarcopenia. *Hum Genet* 2012;**131**:1–31.

238. Carrió E, Suelves M. DNA methylation dynamics in muscle development and

disease. *Front Aging Neurosci* 2015;**7**:19.

239. Jones PA. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012;**13**:484–492.

240. Mendizabal I, Zeng J, Keller TE, Yi S V. Body-hypomethylated human genes harbor extensive intragenic transcriptional activity and are prone to cancer-associated dysregulation. *Nucleic Acids Res* 2017;**45**:4390–4400.

241. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. *Cancer Cell* 2014;**26**:577–590.

242. Jjingo D, Conley AB, Yi S V, Lunyak V V, King I. Jjingo et al. - 2012 - On the presence and role of human gene-body DNA me.pdf. 2012;**3**:462–474.

243. Fujie S, Sato K, Miyamoto-Mikami E, Hasegawa N, Fujita S, Sanada K *et al.* Reduction of Arterial Stiffness by Exercise Training Is Associated with Increasing Plasma Apelin Level in Middle-Aged and Older Adults. *PLoS One* 2014;**9**:e93545.

244. Yamamoto T, Habata Y, Matsumoto Y, Yasuhara Y, Hashimoto T, Hamajyo H *et al.* Apelin-transgenic mice exhibit a resistance against diet-induced obesity by increasing vascular mass and mitochondrial biogenesis in skeletal muscle. *Biochim Biophys Acta - Gen Subj* 2011;**1810**:853–862.

245. Vinel C, Lukjanenko L, Batut A, Deleruyelle S, Pradère JP, Le Gonidec S *et al.* The exerkine apelin reverses age-associated sarcopenia. *Nat Med* 2018;**24**:1360–1371.

246. Boucher J, Masri B, Daviaud D, Gesta S, Guigné C, Mazzucotelli A *et al.* Apelin, a newly identified adipokine up-regulated by insulin and obesity. *Endocrinology* 2005;**146**:1764–71.

247. Chang L, Chiang S-H. Insulin signaling and the regulation of glucose transport. *Mol Med* 2004;**Jul-Dec**:65–71.

248. Tavernarakis N. Ageing and the regulation of protein synthesis: a balancing act? *Trends Cell Biol* 2008;**18**:228–235.

249. Carson JA, Manolagas SC. Effects of sex steroids on bones and muscles: Similarities, parallels, and putative interactions in health and disease. *Bone*

2015;**80**:67–78.

250. Haussler MR, Whitfield GK, Haussler CA, Hsieh J, Thompson PD, Selznick SH *et al.* The Nuclear Vitamin D Receptor: Biological and Molecular Regulatory Properties Revealed. *J Bone Miner Res* 1998;**13**:325–349.

251. Kato S, Takeyama KI, Kitanaka S, Murayama A, Sekine K, Yoshizawa T. In vivo function of VDR in gene expression-VDR knock-out mice. *J Steroid Biochem Mol Biol* 1999;**69**:247–251.

252. Mohammadi Z, Fayyazbakhsh F, Ebrahimi M, Amoli MM, Khashayar P, Dini M *et al.* Association between vitamin D receptor gene polymorphisms (Fok1 and Bsm1) and osteoporosis: A systematic review. *J Diabetes Metab Disord* 2014;**13**:1–9.

253. Ji G-R, Yao M, Sun C-Y, Li Z-H, Han Z. BsmI, TaqI, ApaI and FokI polymorphisms in the vitamin D receptor (VDR) gene and risk of fracture in Caucasians: A meta-analysis. *Bone* 2010;**47**:681–686.

254. Dubińska-Magiera M, Jabłońska J, Saczko J, Kulbacka J, Jagla T, Daczewska M. Contribution of small heat shock proteins to muscle development and function. *FEBS Lett* 2014;**588**:517–530.

255. Rossor AM, Morrow JM, Polke JM, Murphy SM, Houlden H, Laura M *et al.* Pilot phenotype and natural history study of hereditary neuropathies caused by mutations in the HSPB1 gene. *Neuromuscul Disord* 2017;**27**:50–56.

256. Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, Tapscott SJ. Pbx Marks Genes for Activation by MyoD Indicating a Role for a Homeodomain Protein in Establishing Myogenic Potential. *Mol Cell* 2004;**14**:465–477.

257. Maves L, Waskiewicz AJ, Paul B, Cao Y, Tyler A, Moens CB *et al.* Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. *Development* 2007;**134**:3371–3382.

258. McPhee JS, Cameron J, Maden-Wilkinson T, Piasecki M, Yap MH, Jones DA *et al.* The Contributions of Fiber Atrophy, Fiber Loss, in Situ Specific Force, and Voluntary Activation to Weakness in Sarcopenia. *Journals Gerontol - Ser A Biol Sci Med Sci* 2018;**73**:1287–1294.

259. Dallabona C, Diodato D, Kevelam SH, Haack TB, Wong L-J, Salomons GS

et al. Novel (ovario) leukodystrophy related to AARS2 mutations. *Neurology* 2014;**82**:2063–71.

260. Konovalova S, Tyynismaa H. Mitochondrial aminoacyl-tRNA synthetases in human disease. *Mol Genet Metab* 2013;**108**:206–211.

261. Bodine SC, Stitt TN, Gonzalez M, Kline WO, Stover GL, Bauerlein R *et al.* Akt/mTOR pathway is a crucial regulator of skeletal muscle hypertrophy and can prevent muscle atrophy in vivo. *Nat Cell Biol* 2001;**3**:1014–9.

262. Sumter TF, Xian L, Huso T, Koo M, Chang Y-T, Almasri TN *et al.* The High Mobility Group A1 (HMGA1) Transcriptome in Cancer and Development HHS Public Access Conclusion-Further elucidation of HMGA1 function should lead to novel therapeutic strategies for cancer and possibly for other diseases associated with aberrant H. *Curr Mol Med* 2016;**16**:353–393.

263. Huang T, Zhou Y, Zhang J, Wong CC, Li W, Kwan JSH *et al.* SRGAP1, a crucial target of miR-340 and miR-124, functions as a potential oncogene in gastric tumorigenesis. *Oncogene* 2018;**37**:1159–1174.

264. Prosper F, Agirre X, Fernandez AF, Assenov Y, Martin-Subero JI, Balint B *et al.* A DNA methylation fingerprint of 1628 human samples. *Genome Res* 2012;**22**:407–419.

265. Ren Y, Zheng J, Yao X, Weng G, Wu L. Essential role of the cGMP/PKG signaling pathway in regulating the proliferation and survival of human renal carcinoma cells. *Int J Mol Med* 2014;**34**:1430–1438.

266. Tinsley HN, Gary BD, Keeton AB, Lu W, Li Y, Piazza GA. Inhibition of PDE5 by sulindac sulfide selectively induces apoptosis and attenuates oncogenic Wnt/β-catenin-mediated transcription in human breast tumor cells. *Cancer Prev Res* 2011;**4**:1275–1284.

267. Brandan E, Gutierrez J. Role of skeletal muscle proteoglycans during myogenesis. *Matrix Biol* 2013;**32**:289–297.

268. Fadic R, Mezzano V, Alvarez K, Cabrera D, Holmgren J, Brandan E. Increase in decorin and biglycan in Duchenne muscular dystrophy: Role of fibroblasts as cell source of these proteoglycans in the disease. *J Cell Mol Med* 2006;**10**:758–769.

269. Dodds RM, Syddall HE, Cooper R, Benzeval M, Deary IJ, Dennison EM *et*

al. Grip strength across the life course: Normative data from twelve British studies. *PLoS One* 2014;**9**:1–15.

270. Lourenço RA, Pérez-zepeda M, Gutiérrez-robledo L, García-garcía FJ, Rodríguez mañas L. Performance of the European working group on sarcopenia in older people algorithm in screening older adults for muscle mass assessment. *Age Ageing* 2015;**44**:334–338.

271. Yoo J-I, Choi H, Ha Y-C. Mean Hand Grip Strength and Cut-off Value for Sarcopenia in Korean Adults Using KNHANES VI. *J Korean Med Sci* 2017;**32**:868.

272. von Haehling S, Morley JE, Coats AJS, Anker SD. Ethical guidelines for publishing in the journal of cachexia, sarcopenia and muscle: update 2017. *J Cachexia Sarcopenia Muscle* 2017;**8**:1081–1083.

273. Balogun S, Winzenberg T, Wills K, Scott D, Jones G, Callisaya ML *et al.* Prospective associations of low muscle mass and strength with health-related quality of life over 10-year in community-dwelling older adults. *Exp Gerontol* 2019;**118**:65–71.

274. Puthucheary Z, Skipworth JRA, Rawal J, Loosemore M, Van Someren K, Montgomery HE. The ACE gene and human performance: 12 Years on. *Sport Med* 2011;**41**:433–448.

275. Erskine RM, Williams AG, Jones DA, Stewart CE, Degens H. The individual and combined influence of ACE and ACTN3 genotypes on muscle phenotypes before and after strength training. *Scand J Med Sci Sports* 2014;**24**:642–648.

276. Vincent B, De Bock K, Ramaekers M, Van den Eede E, Van Leemputte M, Hespel P *et al.* ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiol Genomics* 2007;**32**:58–63.

277. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat Genet* 2003;**33**:245–254.

278. Tammen SA, Friso S, Choi SW. Epigenetics: The link between nature and nurture. *Mol Aspects Med* 2013;**34**:753–764.

279. Mikeska T, Craig JM. DNA methylation biomarkers: Cancer and beyond.

Genes (Basel) 2014;**5**:821–864.

280. Bharathy N, Ling BMT, Taneja R. Epigenetic Regulation of Skeletal Muscle Development and Differentiation. In: *Epigenetics: Development and Disease*. 2013. pp. 139–150.

281. He L, Khanal P, Morse CI, Williams A, Thomis M. Differentially methylated gene patterns between age-matched sarcopenic and non-sarcopenic women. *J Cachexia Sarcopenia Muscle* 2019;jcsm.12478.

282. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* 1996;**58**:267–288.

283. Burnham KP, Anderson DR. Multimodel Inference. *Sociol Methods Res* 2004;**33**:261–304.

284. Singh P, James RS, Mee CJ, Morozov IY. mRNA levels are buffered upon knockdown of RNA decay and translation factors via adjustment of transcription rates in human HepG2 cells. *RNA Biol* 2019;**0**:1–9.

285. Shekhar A, Lin X, Liu F, Zhang J, Mo H, Bastarache L *et al.* Transcription factor ETV1 is essential for rapid conduction in the heart. *J Clin Invest* 2016;**126**:4444–4459.

286. Arber S, Ladle DR, Lin JH, Frank E, Jessell TM. ETS Gene Er81 Controls the Formation of Functional Connections between Group Ia Sensory Afferents and Motor Neurons. *Cell* 2000;**101**:485–498.

287. Bai SW, Herrera-Abreu MT, Rohn JL, Racine V, Tajadura V, Suryavanshi N *et al.* Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol* 2011;**9**:54.

288. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 2009;**10**:295–304.

289. Tiainen K, Sipilä S, Alen M, Heikkinen E, Kaprio J, Koskenvuo M *et al.* Heritability of maximal isometric muscle strength in older female twins. *J Appl Physiol* 2004;**96**:173–180.

290. Pasquin S, Sharma M, Gauchat JF. Ciliary neurotrophic factor (CNTF): New facets of an old molecule for treating neurodegenerative and metabolic

syndrome pathologies. *Cytokine Growth Factor Rev* 2015;**26**:507–515.

291. Roth SM, Schrager MA, Ferrell RE, Riechman SE, Metter EJ, Lynch NA *et al.* CNTF genotype is associated with muscular strength and quality in humans across the adult age span. *J Appl Physiol* 2001;**90**:1205–10.

292. Consortium Gte, Aguet F, Brown AA, Castel SE, Davis JR, He Y *et al.* Genetic effects on gene expression across human tissues. *Nature* 2017;**550**:204–213.

293. Abney M, McPeek MS, Ober C. Broad and Narrow Heritabilities of Quantitative Traits in a Founder Population. *Am J Hum Genet* 2001;**68**:1302–1307.

294. Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008;**456**:18–21.

295. Matteini AM, Tanaka T, Karasik D, Atzmon G, Chou WC, Eicher JD *et al.* GWAS analysis of handgrip and lower body strength in older adults in the CHARGE consortium. *Aging Cell* 2016;**15**:792–800.

296. Young AI. Solving the missing heritability problem. *PLOS Genet* 2019;**15**:e1008222.

297. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C *et al.* Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Res* 2003;**6**:399–408.

298. Macgregor S, Cornes BK, Martin NG, Visscher PM. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum Genet* 2006;**120**:571–580.

299. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Hum Mol Genet* 2018;**27**:3641–3649.

300. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010;**42**:565–569.

301. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S *et al.*

Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014;**46**:1173–1186.

302. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci* 2012;**109**:1193–1198.

303. Purcell S. Variance Components Models for Gene–Environment Interaction in Twin Analysis. *Twin Res* 2002;**5**:554–571.

304. Solheim TS, Fayers PM, Fladvad T, Tan B, Skorpen F, Fearon K *et al.* Is there a genetic cause of appetite loss?-an explorative study in 1,853 cancer patients. *J Cachexia Sarcopenia Muscle* 2012;**3**:191–198.

305. Lee S-J, Lee Y-S, Zimmers TA, Soleimani A, Matzuk MM, Tsuchida K *et al.* Regulation of Muscle Mass by Follistatin and Activins. *Mol Endocrinol* 2010;**24**:1998–2008.

306. Wiederstein JL, Nolte H, Günther S, Piller T, Baraldo M, Kostin S *et al.* Skeletal Muscle-Specific Methyltransferase METTL21C Trimethylates p97 and Regulates Autophagy-Associated Protein Breakdown. *Cell Rep* 2018;**23**:1342–1356.

307. Grobet L, Pirottin D, Farnir F, Poncelet D, Royo LJ, Brouwers B *et al.* Modulating skeletal muscle mass by postnatal, muscle-specific inactivation of the myostatin gene. *Genesis* 2003;**35**:227–38.

308. Schuelke M, Wagner KR, Stolz LE, Hübner C, Riebel T, Kömen W *et al.* Myostatin Mutation Associated with Gross Muscle Hypertrophy in a Child. *N Engl J Med* 2004;**350**:2682–2688.

309. HITTEL DS, AXELSON M, SARNA N, SHEARER J, HUFFMAN KM, KRAUS WE. Myostatin Decreases with Aerobic Exercise and Associates with Insulin Resistance. *Med Sci Sport Exerc* 2010;**42**:2023–2029.

310. Ast G. How did alternative splicing evolve? *Nat Rev Genet* 2004;**5**:773–782.

311. Brenet F, Moh M, Funk P, Feierstein E, Viale AJ, Socci ND *et al.* DNA Methylation of the First Exon Is Tightly Linked to Transcriptional Silencing. *PLoS One* 2011;**6**:e14524.

312. Li S, Zhang J, Huang S, He X. Genome-wide analysis reveals that exon

methylation facilitates its selective usage in the human transcriptome. *Brief Bioinform* 2018;**19**:754–764.

313. Shayevitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. *Rna* 2018;**24**:1351–1362.

314. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S *et al.* The NCBI BioSystems database. *Nucleic Acids Res* 2010;**38**:D492–D496.

315. Baker-Andresen D, Ratnu VS, Bredy TW. Dynamic DNA methylation: a prime candidate for genomic metaplasticity and behavioral adaptation. *Trends Neurosci* 2013;**36**:3–13.

316. Mericskay M. Wnt5a is required for proper epithelial-mesenchymal interactions in the uterus. *Development* 2004;**131**:2061–2072.

317. Agarwal A, Zhang M, Trembak-Duff I, Unterbarnscheidt T, Radyushkin K, Dibaj P *et al.* Dysregulated Expression of Neuregulin-1 by Cortical Pyramidal Neurons Disrupts Synaptic Plasticity. *Cell Rep* 2014;**8**:1130–1145.

318. Takeuchi T, Duszkiewicz AJ, Morris RGM. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philos Trans R Soc B Biol Sci* 2013;**369**:20130288–20130288.

319. Evans M, Cogan KE, Egan B. Metabolism of ketone bodies during exercise and training: physiological basis for exogenous supplementation. *J Physiol* 2017;**595**:2857–2871.

320. Fukao T, Mitchell G, Sass JO, Hori T, Orii K, Aoyama Y. Ketone body metabolism and its defects. *J Inherit Metab Dis* 2014;**37**:541–551.

321. Saxton RA, Sabatini DM. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* 2017;**168**:960–976.

322. Eliasson J, Elfegoun T, Nilsson J, Köhnke R, Ekblom B, Blomstrand E. Maximal lengthening contractions increase p70 S6 kinase phosphorylation in human skeletal muscle in the absence of nutritional supply. *Am J Physiol Metab* 2006;**291**:E1197–E1205.

323. Song Z, Moore DR, Hodson N, Ward C, Dent JR, O'Leary MF *et al.* Resistance exercise initiates mechanistic target of rapamycin (mTOR) translocation and protein complex co-localisation in human skeletal muscle.

Sci Rep 2017;**7**:5028.

324. Biolo G, Maggi SP, Williams BD, Tipton KD, Wolfe RR. Increased rates of muscle protein turnover and amino acid transport after resistance exercise in humans. *Am J Physiol Metab* 1995;**268**:E514–E520.

325. Burd NA, West DWD, Moore DR, Atherton PJ, Staples AW, Prior T *et al.* Enhanced Amino Acid Sensitivity of Myofibrillar Protein Synthesis Persists for up to 24 h after Resistance Exercise in Young Men. *J Nutr* 2011;**141**:568–573.

326. Woo S-Y, Kim D-H, Jun C-B, Kim Y-M, Haar E Vander, Lee S *et al.* PRR5, a Novel Component of mTOR Complex 2, Regulates Platelet-derived Growth Factor Receptor β Expression and Signaling. *J Biol Chem* 2007;**282**:25604–25612.

327. Hagberg JM, Rankinen T, Loos RJF, Pérusse L, Roth SM, Wolfarth B *et al.* Advances in Exercise, Fitness, and Performance Genomics in 2010. *Med Sci Sport Exerc* 2011;**43**:743–752.

328. Anderson LJ, Liu H, Garcia JM. Sex Differences in Muscle Wasting. In: Mauvais-Jarvis F, editor. *Sex and Gender Factors Affecting Metabolic Homeostasis, Diabetes and Obesity*. Springer International Publishing: Cham; 2017. pp. 153–197.

329. Corrêa Da Costa VM, Moreira DG, Rosenthal D. Thyroid function and aging: Gender-related differences. *J Endocrinol* 2001;**171**:193–198.

330. Horie I, Abiru N, Eto M, Sako A, Akeshima J, Nakao T *et al.* Sex differences in insulin and glucagon responses for glucose homeostasis in young healthy Japanese adults. *J Diabetes Investig* 2018;**9**:1283–1287.

Appendices

Appendix table

Table 1. Summary of genes and pathways associated with muscle mass and strength identified in this thesis

Terms	Categories	Names
Muscle mass	Genes (n = 13)	*ACVR1B, AKT1, FST, H19, IGF1, IGF2, IGF2AS, IGFBP3, IL15-RA, MSTN, MYOG, OPN/SPP1, TGFB1*
	Pathways (n = 10)	Thermogenesis, cGMP-PKG signaling pathway, Aldosterone synthesis and secretion, Cell cycle, Wnt signaling pathway, Focal adhesion, Proteoglycans in cancer, MAPK signaling pathway, Hedgehog signaling pathway, ErbB signaling pathway
Muscle strength	Genes (n = 23)	*APOA1, ATP1A2, CACNA1S, CKM, DNMT3L, ESR1, GR/NR3C1, HIF1A, INS, KBTBD13, LEPR, PPARa, RYR1, VDR, BMP2, FN1, GSC, SMG6, SPP1, TTN, CNTF, RIMS1, ZNF804A*
	Pathways (n = 18)	Phosphatidylinositol signaling system, AMPK signaling pathway, Axon guidance, Oxytocin signaling pathway, Insulin resistance, Pantothenate and CoA biosynthesis, Glucagon signaling pathway, Thyroid hormone signaling pathway, Inositol phosphate metabolism, Human papillomavirus infection, Endocrine and other factor-regulated calcium reabsorption, Gastric acid secretion, Glycerophospholipid metabolism, Vasopressin-regulated water reabsorption, Adrenergic signaling in cardiomyocytes, Insulin secretion, GABAergic synapse, Insulin signaling pathway

Both muscle mass and strength	Pathways (n = 9)	Apelin signaling pathway, Cushing syndrome, Endocytosis, cAMP signaling pathway, Tight junction, Adherens junction, Fc gamma R-mediated phagocytosis, Purine metabolism, Adipocytokine signaling pathway
Muscle injury/recovery	Genes (n = 2)	*CCL2, CCR2*
Methylation regulation	Genes (n = 5)	*MTHFR, BHMT, METTL21C, MTR, MTRR*

Scientific acknowledgements, personal contribution and conflict of interest statements

Scientific acknowledgement

The author would like to acknowledge the participants of the studies, and the co-authors of the papers for their time and efforts. A special thanks to Dr. Praval Khanal for his help in data collecting and genotyping in the MMU project. The author is also grateful to the Erasmus Mundus 'Move-AGE' joint doctorate program and the Research Foundation – Flanders (FWO) for their sponsorship for the doctoral study.

Personal contribution

The author performed secondary data analyses to build genotype predisposition scores in training response and detraining loss of muscle strength in senior volunteers within a KU Leuven based project "Theme 3.6" for paper 1 and 2. During his one year Move-Age mobility period at Manchester Metropolitan University, the author organised, assessed and analysed data of 307 older women within the "Genetics of Sarcopenia" project together with Dr. Praval Khanal in the MMU for paper 3 and 4. The author is responsible for all statistical analyses, writings of the four scientific manuscripts, and the introductory and discussion chapter of this PhD thesis.

Conflict of interest statements

The authors declare no conflicts of interest.