Common genetic variation at the Klotho gene locus and its relationship to age-related phenotypes

A thesis submitted to the degree of Doctor of Philosophy

Hasnat A Amin
Department of Life Sciences
Brunel University London
August 2021

Abstract

The proportion of older adults in Western populations is increasing and there is, therefore, a need to define factors affecting the maintenance of physical and cognitive health in old age. Klotho, an FGF-23 co-receptor, has been shown to increase the lifespan in mice and there are numerous reports of common variants at the Klotho gene locus, particularly those that make up the *KL*-VS haplotype, being associated with age-related phenotypes. However, these reports are based on small sample sizes and are consequently under-powered. The work described in this thesis uses a candidate gene association approach to evaluate the previously reported associations between common genetic variants at the Klotho gene locus and age-related phenotypes in the UK Biobank, a large prospective cohort study of half a million participants. There was preliminary evidence for the following associations: rs2283368 and rs9536338 with longevity; rs495392 with HbA1c; rs141113969, rs2227122 and rs676046 with memory; and rs71436501 and rs78425544 with melanoma survival. However, none of these associations could be adequately replicated and are likely to be unreliable. These results show that the previous reports of associations between Klotho variants and longevity, cognitive function, cardiometabolic disease and cancer are likely to be false positives, which leads to two overall conclusions: one, there is insufficient evidence that common variation at the Klotho gene locus is associated with age-related phenotypes; and two, there is a need for well-powered replication studies.

Declaration

I, Hasnat Amin, declare that this thesis is my own original work unless otherwise acknowledged in the text or by references.

Prof. Cordell generated the genotype counts from the Newcastle 85 Plus study.

Dr Liu and Prof. Wei performed the replication analysis for melanoma survival.

Acknowledgements

I would like to begin by thanking Prof. Alex Blakemore and Dr Fotios Drenos, my supervisors, for their support, both academic and otherwise. Without this support, this thesis would never have been written and the analyses and ideas that it describes would have been immeasurably less developed. I would also like to thank them for making time in their busy schedules, made even more chaotic by the COVID-19 pandemic, to keep me ticking along, to say the least, during said pandemic.

I would also like to thank Dr Terry Dovey, my Research Development Adviser, for his invaluable insights and guidance, particularly with respect to how people outside of my own field view my work.

I would like to acknowledge and thank the College of Health, Medicine and Life Sciences at Brunel University London for providing me with financial support in the form of a studentship, which enabled me to invest my time and energy into research - time and energy which would otherwise have had to be shared with secondary employment.

I would like to recognise the important contribution of the participants who gave, and continue to give, their time and their data, without which the work described in this thesis would never have been possible.

I would like to thank my family for getting me into a position where higher education became a possibility.

Last, but by no means least, I would like to thank a particular individual, who knows who that individual is, for everything that individual has done for me including, but not limited to, providing me with unconditional support throughout my studies.

Table of Contents

Figures and Tables

Glossary of terms

The **Hardy-Weinberg principle** states that genotype and allele frequencies in a population will remain constant from one generation to the next in the absence of other factors (e.g. nonrandom mating). If this is true, then

$$p^2 + 2pq + q^2 = 1,$$

where p is the frequency of one allele at a particular locus and q is the frequency of the other allele at that same locus. Deviations from this equilibrium in a sample of genotyped individuals can indicate genotyping errors.

Imputation, in the context of genetics, refers to predicting the genotype of an individual at a locus that has not been genotyped by using genotyped variants and a reference panel.

Linkage disequilibrium is when two alleles at two different loci are more likely to be inherited together than would be expected by chance.

Mendelian randomisation is a method that uses genetic variants associated with an exposure of interest, but not with any confounders, to assess the causal effect of the genetically predicted exposure on an outcome.

1 Introduction

The work described in this thesis consists of a series of studies carried out to ascertain whether genetic variation at the Klotho gene locus is associated with ageing, as measured through longevity, as well as by traits and diseases associated with old age, including cognitive function, cancer and cardiometabolic diseases. This chapter: summarises the currently-available evidence supporting the idea that genetic factors might play a role in human ageing; and explains the reasons why identifying factors associated with healthy ageing and longevity is becoming ever more important.

1.1 The ageing population

In the UK, much as in the rest of the developed world, the proportion of older adults is increasing. In 2009, 16.2% of the UK population was aged 65 years or more and by 2039, this is expected to increase to 23.9%. At the same time, due to declining birth rates, the proportion of the population aged under 16 years is expected to decrease: in 2009, individuals aged under 16 years made up 18.9% of the population; and by 2039, this is expected to decrease to 16.9% [1] - please see figure 1.1 for an additional illustration of this point [2]. This will put further strain on our health systems and is very likely to also affect economic growth. Generally, as people grow older, the incidence of several diseases tends to increase. These include, but are not limited to, dementia, cardiovascular disease and cancer: please see figures 1.2 and 1.3 [3]. In addition to the human suffering experienced by the patients and their families and friends, there are economic consequences: healthcare costs rise because the prevalence of age-related diseases increases (because there are more older people); and there are proportionately fewer people in the population who are able to work and/or care for the elderly, so it becomes increasingly difficult to meet the aforementioned increases in healthcare costs.

Figure 1.1 UK population by age group

Population estimates, Principal population projections, 2016-based, Office for National Statistics [2].

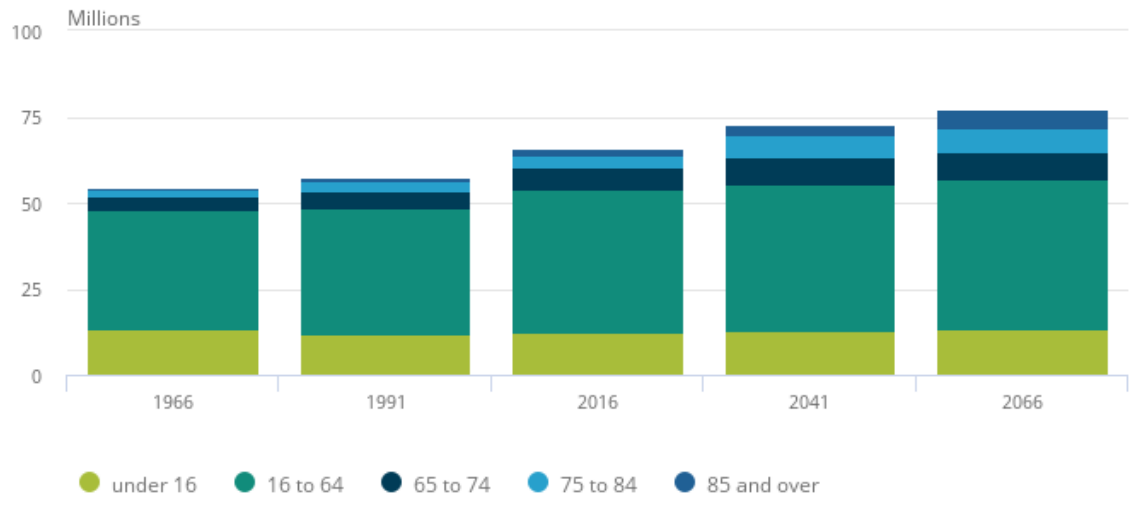


Figure 1.2 Incidence of age-related diseases in the UK (men)

Incidence per 100,000 of neoplasms, ischaemic heart disease, stroke, type 2 diabetes and dementia in men [3].

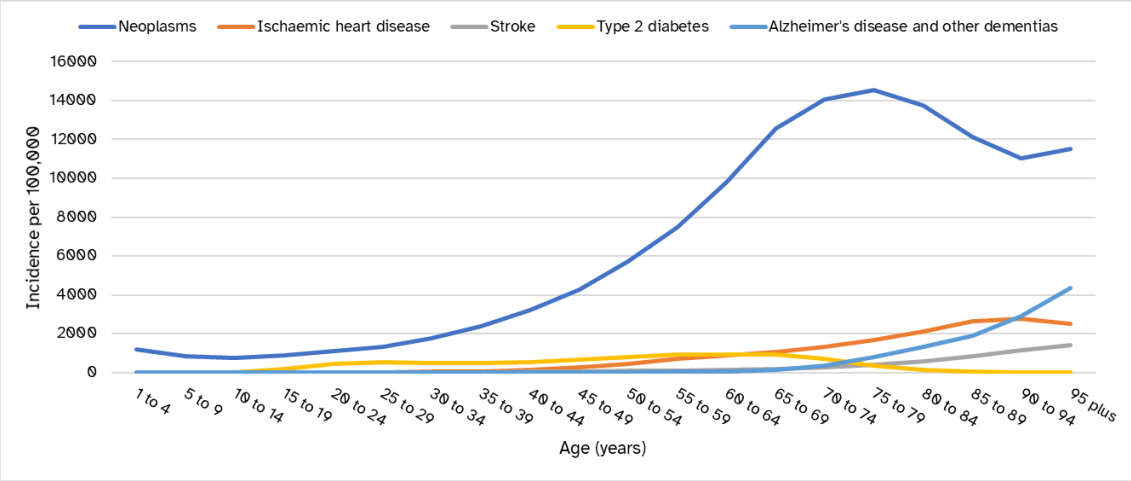
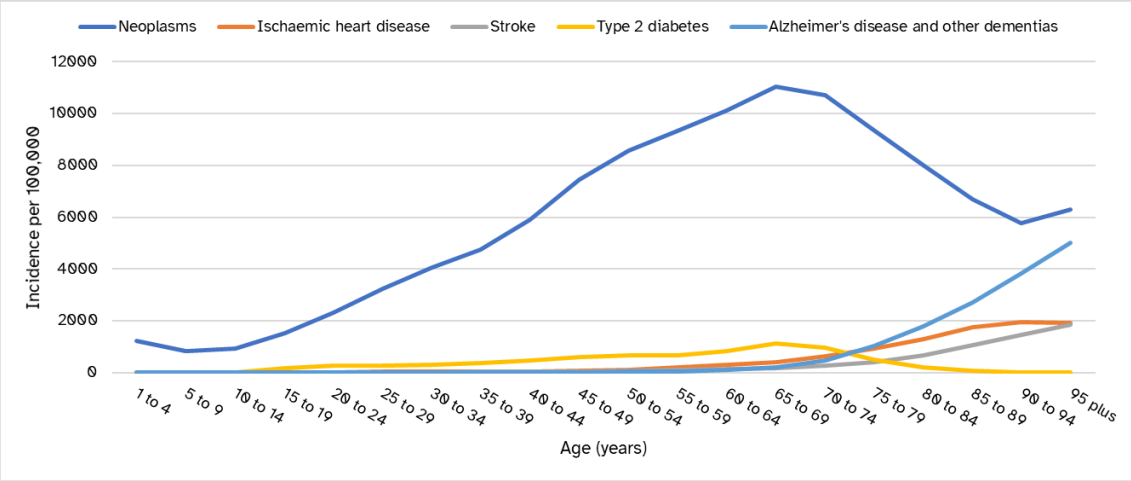


Figure 1.3 Incidence of age-related diseases in the UK (women)

Incidence per 100,000 of neoplasms, ischaemic heart disease, stroke, type 2 diabetes and dementia in women [3].



It is also important to note that many older people are still very much able to contribute economically both through paid employment and voluntary work (including unpaid childcare) and in 2017, the value of these contributions was estimated to be about £160 billion [4]. However, older people can only work if they are physically and mentally healthy enough to do so.

1.2 Measures of ageing in humans

One approach to mitigate the predicted economic and social costs of ageing is to identify factors that affect how well people age. To do this, one would need one or more universal measures of ageing and a possible starting point could be to determine some so-called hallmarks of ageing (Figure 1.4), proposed by López-Otín *et al.* [5]. Briefly:

- Genomic instability refers to the accumulation of DNA damage and the declining ability of an ageing organism to be able to repair this damage. There are many lines of evidence that suggest that genomic instability plays a role in ageing, e.g. mutations in the WRN gene that encodes a DNA repair protein cause Werner's syndrome (a progeroid syndrome where patients manifest the same symptoms that would be expected as one gets older) [6] and increased expression of BubR1, an enzyme that contributes to accurate chromosome segregation [7], extends healthy lifespan in mice [8].
- Telomere attrition is where the regions at the end of chromosomes, called telomeres, become progressively shorter with each cell cycle because DNA polymerases lack the ability to replicate the ends of linear DNA molecules and the cells lack telomerase (an enzyme whose function it is to do exactly this). Telomerase shortening and damage are known causes of cellular senescence (another hallmark) and ageing [9].
- Epigenetics, in simple terms, is how cells regulate the expression of genes and genomic regions without alterations to the DNA sequence itself. These mechanisms include DNA methylation, post-translational modification of histones and chromatin remodelling and there is evidence of epigenetics playing a role in ageing and longevity across multiple organisms and it is thought that the dysregulation of gene expression that is likely to occur as a result of these epigenetic changes contributes to the other hallmarks described here [10].

- Proteostasis refers to various molecules and pathways within cells that prevent proteins from misfolding or aggregating and promote the clearance of damaged or dysfunctional proteins. As organisms age, the ability of their cells to maintain proteostasis decreases. A good example of how a loss of proteostasis leads to age-related disease is Alzheimer's disease [11].
- Nutrient sensing refers to how anabolic signalling (i.e. the somatotroph growth hormone/insulin-like growth factor (GH/IGF1) axis and downstream effectors such as mTOR) and nutrient sensing proteins such as AMPK and sirtuins affect ageing in organisms. Generally speaking, the evidence suggests that anabolic signalling accelerates ageing, but upregulation of proteins such as AMPK and SIRT1 (i.e. proteins which are upregulated in low-energy states that occur during, for example, caloric restriction) appears to be associated with longevity [12].
- Mitochondria are responsible for generating energy in the form of ATP and can be considered as cells within cells that have their own quality control mechanisms. However, over time, mitochondria become damaged and are less able to repair that damage, much like the cells in which they reside. This in turn disrupts and damages their host cells and contributes to the loss of proteostasis, for example [13].
- Cellular senescence is when a cell is no longer able to divide due to a variety of factors including, but not limited to, telomere shortening and DNA damage (which are both hallmarks of ageing themselves). Whilst cellular senescence has protective effects (e.g. tumour suppression), there is evidence that suggests the ability of an ageing organism to replace senescent cells declines with age and these senescent cells create a pro-inflammatory environment (sometimes referred to as the senescence-associated secretory phenotype) that drives disease processes associated with ageing [14].
- Stem cells are cells that are able to differentiate into many different specialised cell types. The body has pools of stem cells that are important for the replacement of damaged cells (e.g. red and white blood cells) and tissues. However, as the organism ages, these pools of stem cells become smaller, which contributes to various age-related diseases and phenotypes, such as a weakened immune system and muscle loss. The reasons for a reduction in these stem cell pools are likely to be caused by both an accumulation of damage (e.g. DNA

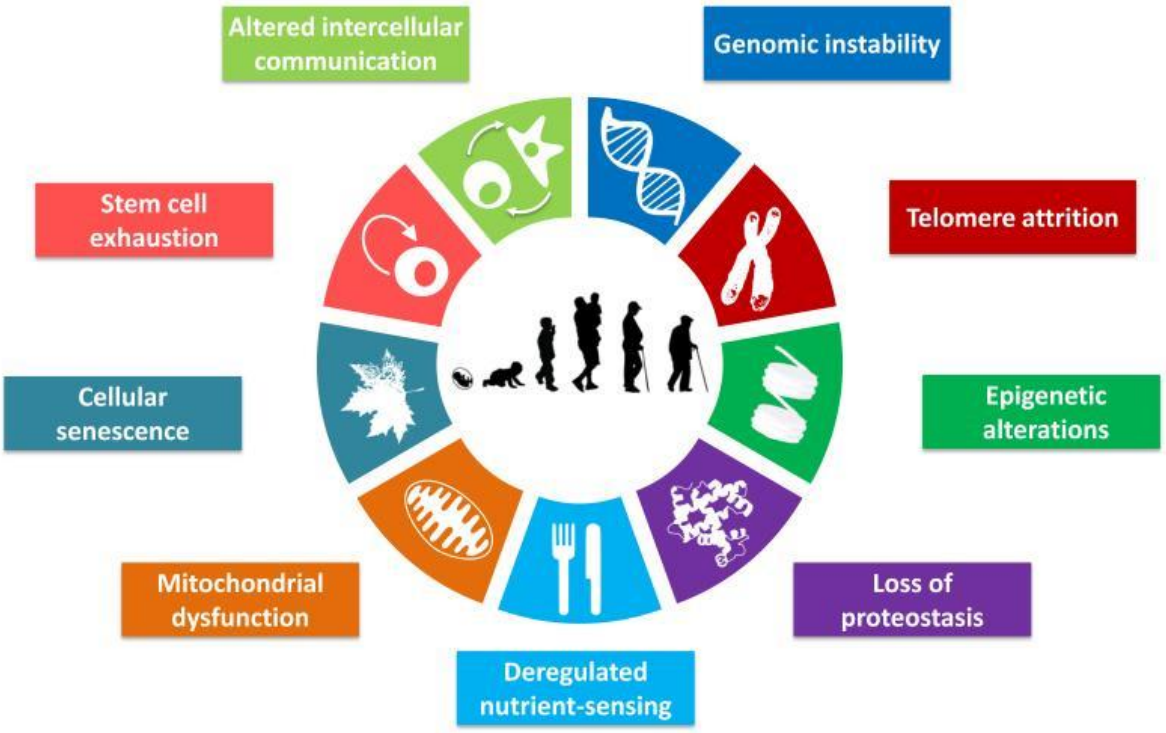
damage and loss of proteostasis) as well as an extracellular environment that is pro-inflammatory and pro-senescent [15].

- In addition to the changes within cells, there are changes in the extracellular environment and intercellular communication. For example, the presence of senescent cells promote senescence in neighbouring cells both through the senescence-associated secretory phenotype and through processes involving ROS and gap junction-mediated cell-cell contacts [16].

However: these hallmarks of ageing are difficult to measure in humans on a large scale. Therefore, studies undertaken to understand factors that affect human ageing often use proxy measures and these will be discussed in this section.

Figure 1.4 The hallmarks of ageing

The hallmarks of ageing. Obtained from López-Otín *et al.* [5].



Lifespan is sometimes used as a measure of ageing. However, the pursuit of longevity (which can be defined as above-average lifespan) as an aim in and of itself for the purposes of mitigating the social and economic impact of ageing assumes that people who live longer also spend less time in a state of ill health - there is published evidence to support this assumption. For example: a 2018 review of available epidemiological evidence by Qiu *et al.* [17] suggests that, despite the marked increase of dementia with age, dementia is not an inevitable fact for those who live past 100 years; and Pavlidis *et al.* [18] found that cancer incidence reduces significantly after 90 years of age and that cancers that occur in people who are extremely old are less likely to be life-threatening. There is also evidence that centenarians experience morbidity at later age [19]. Together, this suggests that studying long-lived individuals is a valid approach to identifying factors that affect how well people age and this was one of the reasons for use of Newcastle 85 Plus study data for the work described in Chapter 3.

At present, however, there are relatively few individuals who have reached ages of 100 years or more. Therefore, it is useful to look at other possible markers of ageing, such as: healthspan, which is often defined as the number of years an individual is free from age-related diseases; clinically-measured biomarkers, such as lipid levels and markers of inflammation, kidney and liver function, and indices that combine these measures; and phenotypes that measure the functional ability of a person, including cognitive function, gait speed and measures of frailty. The work described in chapters 4, 5 and 6 explores the association of Klotho variants with cardiovascular disease and its risk factors, measures of cognitive function and cancer risk and survival, respectively.

1.3 Genetics of human ageing

Human ageing is affected by both genetic and environmental factors and their interaction. Here, the evidence pertaining to genetics being a contributory factor in human ageing is summarised.

If genetics plays a role in ageing, it follows that at least some portion of the variation seen in lifespan and susceptibility to age-related disease between individuals must be heritable. Studies using twins and families with a centenarian have yielded estimates that over 20% of the variation seen in lifespan is heritable [20,21]. Furthermore, studies

of age-related diseases such as type 2 diabetes, dementia and cardiovascular disease have produced heritability estimates in excess of 40% [22-24].

Further evidence of genetics influencing the ageing process comes from the existence of progeroid syndromes inherited in a classical Mendelian fashion. Two well-characterised progeroid syndromes are Hutchinson-Gilford progeria syndrome and Werner syndrome, which are caused by mutations in the LMNA and WRN genes, respectively, and these syndromes manifest the same symptoms that would be expected as one gets older [6].

In addition, there is a large body of evidence from model organisms such as roundworms [25], fruit flies [26] and mice [27] that implicates a wide variety of different genes as being involved in the ageing process. Klotho, the subject of the work described in this thesis, was also first discovered in a mouse model (see Section 1.5.1 for description).

1.4.1 Candidate gene association studies

The aforementioned studies inspired numerous candidate gene association (CGA) studies in humans. The purpose of these CGA studies was to examine whether genetic variants located within and/or near genes previously implicated in ageing in model organisms are associated with longevity or other proxy measures of ageing. Numerous authors have used this CGA approach to obtain evidence for an association between Klotho variants and age-related phenotypes and this evidence will be reviewed in the relevant chapters.

However, despite a large number of these CGA studies suggesting that Klotho variants are associated with age-related phenotypes, there are limitations, some specific to CGA studies and others that are perhaps applicable to other approaches and even fields. Association studies of Klotho have been carried out in a variety of age ranges and population groups and all use relatively small sample sizes, so one cannot be sure that the finding will be applicable to the general population.

In addition to the limitation specific to CGA studies, it could be argued that there is a replication problem in genetics [28]. This has two causes: first, since CGA studies are hypothesis-driven, they may be based on a body of literature that contains insufficient contributions that seek to replicate previous findings and is therefore biased; and second, there are almost no attempts to replicate CGA studies in large sample sizes using a CGA approach [29]. These two factors, in combination with publication bias (which favours positive findings), mean that there is an urgent need to confirm the results of even widely-reported association studies, including of the *KL* gene.

1.4.2 Genome-wide association studies

A genome-wide association study (GWAS) uses a hypothesis-free approach and examines (up to) millions of common variants across the genome for an association with a phenotype of interest. Whilst this approach does not suffer from hypotheses formed from evidence that is subject to publication bias, it does have drawbacks. Firstly, a GWAS involves around a million independent statistical tests and subsequent correction for these tests (see Section 2.8) means that most associations are assumed to be a chance finding and are not subject to downstream analyses. Secondly, due to the scale of a GWAS, only an additive genetic model is typically considered (see Section 2.9) and the possibility of variants having different effects in different sub-populations is typically not tested. CGA studies are much smaller, which lowers the multiple testing burden and allows for more extensive downstream analyses. These limitations warrant the use of the CGA approach to investigate the relationship between common *KL* variants and age-related phenotypes.

Although not available at the start of this project, two GWASs of parental lifespan [29,30], published in 2019, do provide some interesting insight into the genetics associated with ageing. Firstly, *KL* is not identified as a locus of interest, but this may be due to the limitations described in the above paragraph. Secondly, the genes located on or near the statistically significant associations appear to be associated with age-related diseases (see Table 1.1); this suggests that the associations between variants and lifespan are driven by associations between these variants and age-related disease.

Table 1.1 Genes mapped to variants associated with parental lifespan

Variants associated with parental lifespan in both Timmers *et al.* and Wright *et al.* [29,30]. Adapted from Melzer *et al.* [31].

Variant	Mapped genes	Gene name	Variant position	Associated disease
rs429358	APOE	Apolipoprotein E	Missense	Cardiometabolic, dementia
rs10455872	LPA	Lipoprotein A	Intronic	Cardiometabolic
rs8042849	CHRNA3/5	Cholinergic receptor nicotinic α3/5 subunit	Intronic	Smoking related
rs142158911	LDLR	Low-density lipoprotein receptor	Intergenic	Cardiometabolic
rs11065979	SH2B3, ATXN2	SH2B adaptor protein 3, ataxin 2	Intergenic	Cardiometabolic, cancer, autoimmunity
rs1556516	CDKN2B-AS1	CDKN2B antisense RNA 1	Intronic	Cardiometabolic, cancer

1.5 Klotho

1.5.1 Discovery of Klotho

The first evidence of Klotho potentially playing a role in ageing was uncovered serendipitously by Kuro-o *et al*. [32], who found that disruption of a gene in mice (which the authors named Klotho, or *kl*, after Κλωθώ, who is one of the Three Fates in Greek mythology that is said to spin the thread of life) resulted in a shortened lifespan and other age-related phenotypes (e.g. sarcopenia, thin skin, osteoporosis and vascular calcification), but that these *kl* deficient mice did not differ substantially from wildtype mice during their first few weeks of life, suggesting a role in ageing process.

It was later discovered that the *fgf23*-knockout mouse has a similar phenotype to the aforementioned *kl* deficient mouse, suggesting that perhaps the two proteins interacted in some way to perform a function. This prompted further studies that provided evidence that Klotho was a co-receptor for FGF23 (fibroblast growth factor 23, which is a bone-derived hormone that suppresses phosphate reabsorption and vitamin D hormone synthesis in the kidney) [33,34]. Phosphate levels have been associated with lifespan in fruit flies and in mammals [35,36]. There is an 86% homology between mouse and human forms of Klotho protein [37].

1.5.2 Klotho is an FGF23 co-receptor

In humans, the Klotho gene is located on chromosome 13 (position 33,590,571 to 33,640,282 using build GRCh37.p13) and encodes an FGF-23 co-receptor (see below). It has five exons (see Figure 1.5) [38]. The full-length Klotho protein consists of two domains, *KL*1 and *KL*2. There are two truncated forms of Klotho, referred to as soluble Klotho, one of which is produced by enzymatic cleavage and the other is produced via alternative splicing (see Figure 1.6) [39].

Figure 1.5 Klotho gene structure

Structure of the Klotho gene with positions of rs1207568, rs9536314 and rs564481 shown. Obtained from Riancho *et al.* [38].

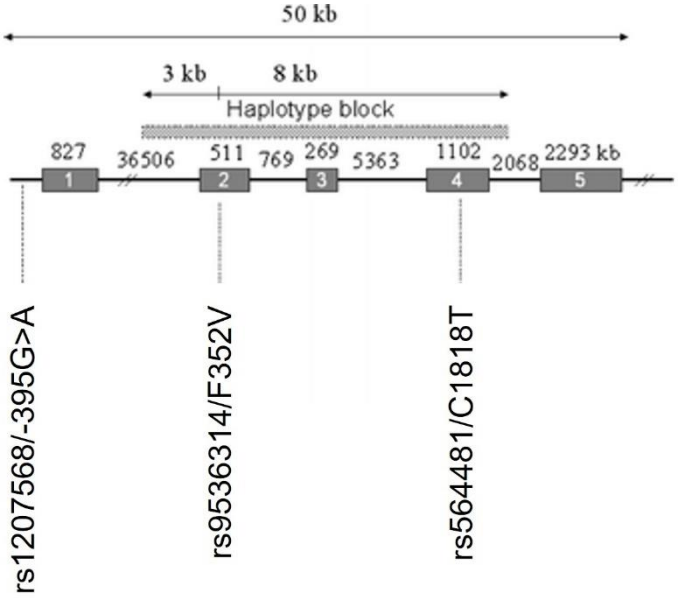
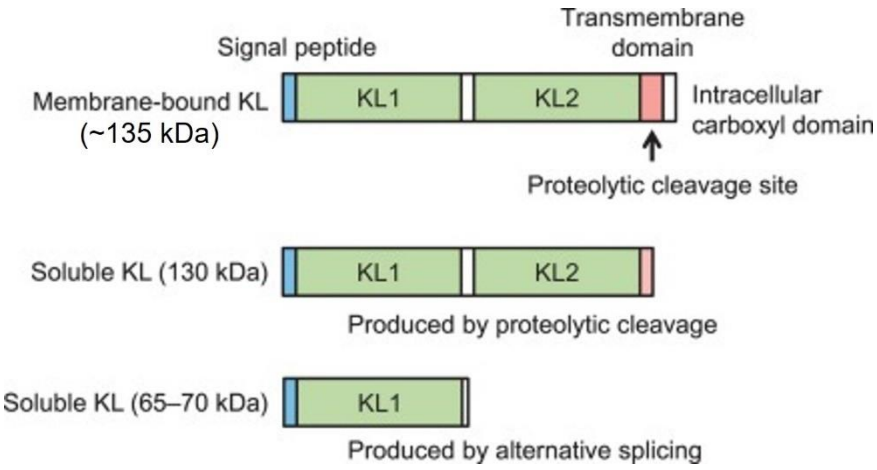


Figure 1.6 Klotho protein domains

Klotho protein domains. Adapted from Kinoshita and Kawai [39].

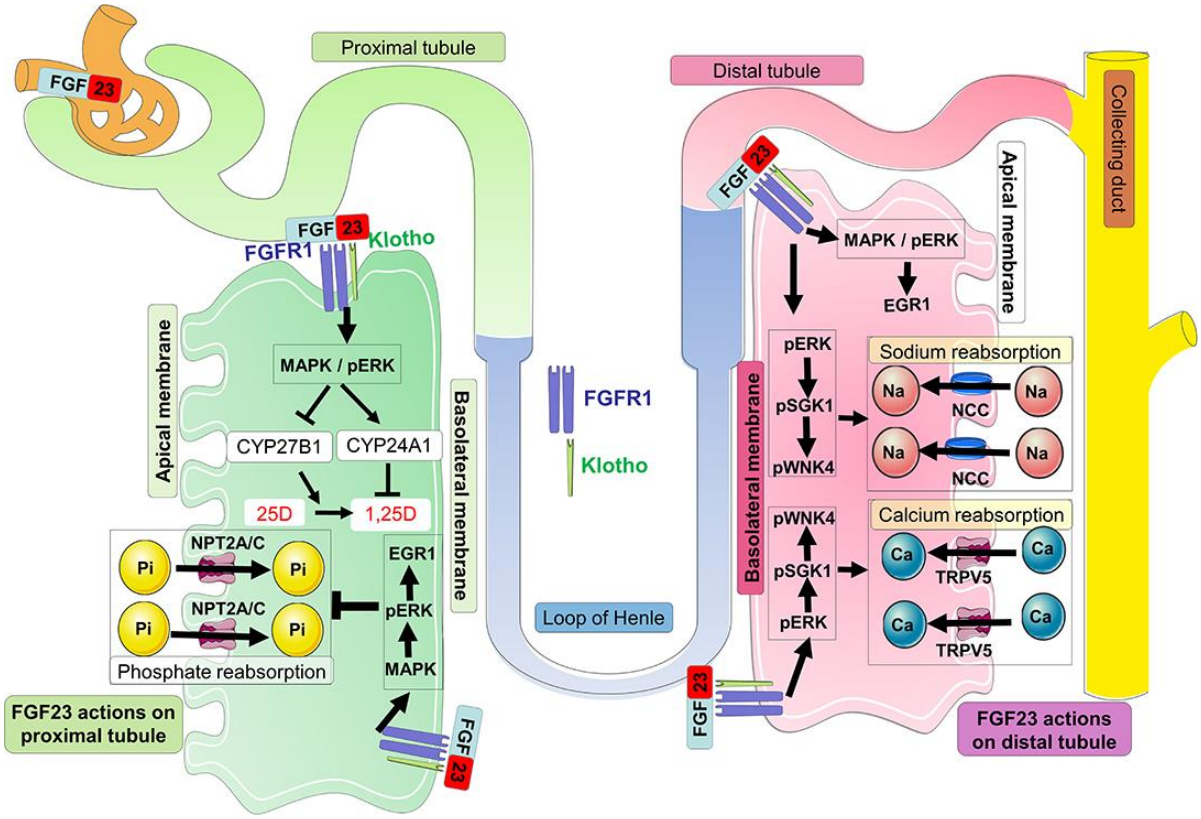


Klotho is an FGF23 co-receptor. FGF23 has a weak binding affinity for FGFR1c (fibroblast growth factor receptor 1c isoforms). Klotho acts to stabilise the binding between FGF23 and FGFR1c. These FGF23–FGFR1c–Klotho complexes then form dimers dependent on heparan sulphate (a polysaccharide found in animals that binds to a wide array of extracellular proteins) and allow FGF23 signalling to take place [40].

There are several downstream effects of FGF23 signalling in the kidney - please see figure 1.7 [41]. Activation of MAPK-ERK signalling results in suppression/internalisation of NPT2A/C sodium phosphate co-transporters [42,43], which in turn reduces phosphate reabsorption [44]. In addition, CYP27B1 (an enzyme that converts vitamin D into its active form, 1,25-dihydroxyvitamin D_3) is suppressed [45] and CYP24A1 (an enzyme that catabolises vitamin D) is upregulated [46]. Since vitamin D is involved in dietary phosphate uptake and in phosphate reabsorption (via PTH), the overall effect of Klotho-dependent FGF23 signalling is to reduce phosphate concentrations in the blood. In addition, FGF23 signalling promotes the reabsorption of sodium chloride by regulating the abundance of NCC at the membrane [47] and also promotes the reabsorption of calcium due to an increased abundance of TRPV5 (the epithelial calcium channel Transient Receptor Potential Vanilloid-5) at the membrane [48].

Figure 1.7 FGF23 signalling in the kidney

FGF23 signalling in the kidney. Obtained from Agoro *et al.* [41].



It should also be noted that there is another form of Klotho, referred to as soluble Klotho, which is generated when the extracellular domains of the full-length Klotho protein are cleaved by ADAM10/17 proteases [49]. However, the function of this form is unclear.

1.5.3 Common variants at the Klotho gene locus

Several common genetic variants at the Klotho gene locus have been identified (the location of some of these is shown in Figure 1.5). The evidence for their potential functional consequences on Klotho function is summarised in this section and the evidence for their association with age-related phenotypes is reviewed elsewhere.

KL-VS is the haplotype that is used to refer to two common variants, rs9536314 and rs9527025, that are in complete linkage disequilibrium with each other. Linkage disequilibrium is when two alleles at two different loci are more likely to be inherited together than would be expected by chance. Data from the LDlink database [50] shows that there are no individuals who carry the minor allele for one variant without carrying the minor allele for the other.

The rs9536314 variant results in a phenylalanine to valine substitution at position 352 and the rs9527025 variant results in a cysteine to serine substitution at position 370 [51]. Based on previously published evidence, the *KL*-VS variants appear to affect FGF23 signalling [52] but also the amount of soluble Klotho in the blood and in the cerebrospinal fluid [53-55], though the evidence for the effect of *KL*-VS on soluble Klotho concentration is contradictory [56].

G395A (rs1207568) is a promoter variant (MAF = 0.16). Carrying the A allele is reportedly associated with increased Klotho expression based upon a luciferase reporter assay [57]. However, a study of human vascular tissue samples revealed that carrying the A allele reduces Klotho expression [58], so the direction of effect is still unclear.

1.6 Aims and Objectives

Much of the evidence for associations between common Klotho variants and age-related phenotypes described by previous authors is contradictory and based on relatively small sample sizes. Positive results, reviewed in subsequent chapters, have encouraged researchers to prioritise Klotho as a target for anti-ageing therapies. As described above, however, there have been methodological limitations in previous CGA studies (including poor power). Therefore, it is necessary to explore these associations in a larger sample size. The work described in this thesis was predominately carried using data from the UK Biobank (UKB), which is a prospective cohort study consisting of over 500,000 participants recruited between 2006 and 2010 from across the UK (see Section 2.2 for further details).

The main aim of this work was to use the candidate gene association approach in the UK Biobank, leveraging the greatly increased power to evaluate whether or not Klotho is a solid candidate for further investigation in anti-ageing research. To achieve this aim, the following steps were undertaken:

- the genetic and phenotypic data available as part of the UKB study were used to attempt to replicate, in British Caucasians, previously-reported associations between the Klotho *KL*-VS haplotype and longevity, cognition, cancer and cardiometabolic phenotypes;
- the CGA approach was then used to search for novel associations between other variants, both directly-genotyped and imputed, at the Klotho gene locus and phenotypes related to the four areas above;
- and, where suggestive evidence was found for novel associations, explorations of other datasets and resources were undertaken (*in silico*) to generate hypotheses of how any newly-associated variants might affect the relevant phenotype(s).

2 Participants and Methods

2.1 Introduction

The aim of this chapter is to introduce the UK Biobank (UKB) resource and to describe the statistical methods that will be referred to in subsequent chapters. If a particular aspect of the UKB data and/or a particular statistical model is relevant only to the work in a specific chapter, it will be described within that chapter.

2.2 The UK Biobank

The UK Biobank (UKB) is a large prospective cohort study consisting of over 500,000 participants recruited from across the UK between 2006 and 2010. The participants were aged between 40 and 69 years at the time of recruitment and were invited to one of 22 assessment centres where they provided blood, urine and saliva samples and underwent physical measurements, touchscreen questionnaires and tasks and verbal interviews. A subset of the participants (about 25,000) were invited for reassessment and approximately 100,000 participants have undergone or will undergo MRI (magnetic resonance imaging), ultrasound and DXA (dual-energy X-ray absorptiometry) imaging. Genetic data is available in around 488,000 participants. The ethical approvals and quality control (QC) of the genetic data are detailed in this chapter. The phenotypes used and how they were prepared for analysis will be described in the relevant chapter.

In addition, participants have been, and will continue to be, followed up by linkage to national cancer and death registries and to NHS health records. UKB provides summarised data consisting of diagnoses of each participant and the corresponding dates on which they were diagnosed. In addition, UKB provides additional information on participants who have been diagnosed with cancer, e.g. histology data.

2.3 Ethics Approval

UKB has ethical approval from the North West Multi-centre Research Ethics Committee (16/NW/0274). The work described in this thesis was carried out under UKB application

19968 and was approved by the College of Health and Life Sciences Research Ethics Committee on 29 November 2017 (7689-LR-Nov/2017- 8873-1).

2.4 UK Biobank genetic data

The UKB genetic data contains genotypes for 488,377 participants. The term genotype will be used in this thesis to describe the combination of alleles that a given participant carries at a particular genetic locus.

DNA was extracted from blood samples obtained from the participants during their visit to a UKB assessment centre. Affymetrix Research Services Laboratory genotyped 49,950 samples at 807,411 markers in 11 batches using the UK BiLEVE (UK Biobank Lung Exome Variant Evaluation) Axiom Array [59] and genotyped 438,427 samples at 825,927 markers in 95 batches using the UK Biobank Axiom Array. After an initial QC by Affymetrix [60], genotypes for 488,377 unique participants at 812,428 variants were used by UKB to generate QC metrics [61].

2.4.1 Summary of UKB marker-based quality control metrics

UKB identified poorly-genotyped variants that are not suitable for analyses by comparing the genotype frequencies between experimental factors (i.e. batch effects, plate effects, sex effects, and array effects). If a genetic variant has been genotyped correctly, then the genotype frequencies of each variant should remain consistent across batches, plates and genotyping arrays, and the genotype frequencies of each variant located on the autosomes should remain consistent between males and females.

Before performing any QC on the genetic variants, participants who did not have European ancestry were removed. This is because genotype frequencies at a particular genetic variant can differ between individuals of different ethnic ancestries. Fisher's exact test was used to compare genotype counts for each genetic variant in each batch with genotype counts for the same genetic variant in the whole sample (i.e. all batches). A p-value of 10^{-12} or smaller was considered to be sufficient evidence to reject the null hypothesis that the genotype counts at a particular genetic variant in a given batch were not different from the genotype counts across all batches for that genetic variant. The

same approach was used to compare genotype counts between plates, genotyping arrays, and sexes.

In addition to the aforementioned tests, UKB identified genetic variants whose allele frequency distributions differed significantly ($p \leq 10^{-12}$) from Hardy-Weinberg equilibrium, a principle that states that genotype and allele frequencies in a population will remain constant from generation to generation in the absence of factors such as non-random mating, migration and mortality and, in the case of experiments such as the UKB, genotyping errors. Furthermore, control samples from the CEU (Utah Residents with Northern and Western European Ancestry) group of the 1000 Genomes project were included in each plate. It is expected that the genotype at each variant should be the same for each control sample replicate. The genotypes from these control samples were, therefore, used to compute a discordance metric, *d*, and variants with a $d \geq 0.05$ were not part of the released dataset.

If a variant failed one or more of the aforementioned tests in a given batch, genotype calls for that variant were set to missing for participants genotyped in that batch in the data release. UKB provided a text file listing each genotyped variant and whether or not that variant failed one or more QC tests in any of the batches.

2.4.2 UKB sample-based quality control metrics and analyses

It is necessary to remove samples (i.e. individuals) if they are likely to have been genotyped poorly. Then related pairs of participants must be identified and a quantitative measure of the ancestral background of each participant needs to be generated based on the genetic data.

To carry out the QC and analyses to achieve the aforementioned aims, only SNPs that were genotyped on both arrays and passed QC across all batches were used by UKB. Heterozygosity (the number of heterozygous genotypes an individual has divided by the total number of genotyped variants) and sample-level missingness are used to identify samples that are of poor quality. However, these metrics are affected by the ancestry of an individual and the UKB is an ancestrally heterogenous sample. Principal components analysis (PCA) was carried out on a subset of unrelated individuals (using KING) [62] to

provide a quantitative measure of the ancestral background of each participant, and the heterozygosity and missingness values for each participant were adjusted for their ancestry. These adjusted values were used to identify participants who were outliers for heterozygosity and missingness and were flagged in the sample QC file provided by UKB.

UKB also flags individuals who are likely to have sex chromosome aneuploidies. This was done by comparing the intensities of markers on the Y and X chromosomes.

UKB also provide a list of related pairs, generated using KING software [62]. The SNPs selected for inferring relatedness between samples had small loads on the first 3 genetic PCs to minimise relatedness estimates being artificially high due to admixture (i.e. the presence of DNA from a distantly related ethnic group). Furthermore, individuals flagged as outliers for heterozygosity and missingness were also excluded.

The PCA of the genetic data was repeated, this time excluding poorly-genotyped individuals as well as related pairs. The first 40 PCs are provided by UKB as part of the data release, which are used to adjust for population stratification that may be present within one ethnic group. The White British subset was defined by UKB as the intersect of self-reported ethnic background and the results of the PCA.

2.4.3 Imputation

SHAPEIT3 [63] was used by UKB to estimate haplotypes (phasing), i.e. determining which allele is on which chromosome, using variants that passed QC for all batches. The accuracy of the phasing was assessed by using the mother-father-child trios in the UKB.

A modified version of the IMPUTE2 program [64] was used by UKB to impute (i.e. predict the genotype of an individual at a locus that has not been genotyped by using genotyped variants and a reference panel) about 96 million variants. The genotypes at some variants are predicted more reliably than others and UKB provides an information score for each imputed variant as a measure of this reliability. A variant imputed in a sample of 100,000 individuals with an information score of 0.8 is the same as that variant being perfectly genotyped in a sample of 80,000 individuals.

The data release provides a set of three probabilities per variant for each individual, and each probability represents the likelihood that the individual has a particular genotype at a particular variant. These genotypes are converted into dosages. For example, say that an individual can have one of three genotypes at a particular variant (TT, GT, and GG) and that the probability of that individual having these genotypes are 0.01, 0.97, and 0.02. The genotypes would ordinarily be coded as 0, 1 and 2 (see Section 2.8). However, to take the uncertainty of imputation into account, a dosage is calculated by weighting the three genotypes by their respective probabilities and obtaining a mean. This dosage is used as a predictor variable (see Section 2.7). In this example, the dosage is:

$$\frac{(0.01 \times 0) + (0.97 \times 1) + (0.02 \times 2)}{0.01 + 0.97 + 0.02} = 1.01$$

2.5 Selection of variants for analysis

Variants that failed the previously-described QC tests for any batch that they were genotyped in were excluded. In addition, variants with a missingness of >1% and/or a minor allele frequency (MAF) of <0.01 were also excluded. For imputed variants, all variants with an INFO score of <0.8 were excluded. The *KL* gene is located at position 33,590,571 to 33,640,282 on chromosome 13 (GRCh37.p13) and 247 variants passed QC within ±5 Kb of *KL*. These were selected for the subsequent analyses.

In chapters 3, 4 and 6, due to a small number of cases, variants with a MAF of <0.05 were excluded. It should also be noted that *KL*-VS is always tested separately. For these reasons, the number of variants being tested in each chapter may be less than 247.

2.6 Selection of participants for analysis

Samples were excluded from the analysis if they were determined to be outliers for missingness and/or heterozygosity and/or if they had any sex chromosome aneuploidies as well as if the genetically inferred sex differed from the reported sex. Samples which did not have a genetically-determined White British ancestry were also excluded. A list

of related individuals was also provided by UK Biobank and one individual from each related pair was excluded at random.

2.7 Statistical Methods

Multiple linear regression

A multiple linear regression allows one to estimate how many units the continuous response variable changes by for each unit increase in a predictor variable (i.e. beta) assuming that all other predictor variables are kept constant. The method assumes the following: the relationship between each predictor variable and the response variable is linear; the residuals (i.e. the difference between the actual value of the response variable and the predicted value of the response variable) are consistent for all values of the predictor variable (homoscedasticity); the predictor variables are normally distributed - this does not make much of a difference at large sample sizes; [65] and the predictor variables are not strongly correlated to each other.

$$\begin{aligned}\text{phenotype} = \beta_0 + \beta_1\text{genotype} + \beta_2\text{age} + \beta_3\text{sex} + \beta_4\text{PC1} + \beta_5\text{PC2} + \beta_6\text{PC3} + \beta_7\text{PC4} \\ + \beta_8\text{chip} + \varepsilon\end{aligned}$$

Multiple logistic regression

A multiple logistic regression is used when the response variable is binary (i.e. case-control). It allows one to estimate the change in log odds (i.e. ln(OR), where OR = p/(1-p), where p = probability of observing a case) for each unit increase in a predictor variable assuming that all other predictor variables are kept constant. The assumptions made in a logistic regression are similar to those made in a linear regression.

$$\begin{aligned}\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\text{genotype} + \beta_2\text{age} + \beta_3\text{sex} + \beta_4\text{PC1} + \beta_5\text{PC2} + \beta_6\text{PC3} + \beta_7\text{PC4} \\ + \beta_8\text{chip}\end{aligned}$$

COX PROPORTIONAL HAZARDS MODELS

The Cox proportional hazards model is used to estimate the effect of a given predictor on the rate of a particular event happening at a particular point in time. This effect is known as a hazard ratio. The model assumes that the hazard ratio is constant over time.

$$H(t) = H_0(t) + e^{(\beta_0 + \beta_1 \text{genotype} + \beta_2 \text{age} + \beta_3 \text{sex} + \beta_4 \text{PC1} + \beta_5 \text{PC2} + \beta_6 \text{PC3} + \beta_7 \text{PC4} + \beta_8 \text{chip})}$$

2.8 Multiple testing

The statistical models used throughout this thesis produce a p-value, which is the probability of observing a test statistic that is equally or more extreme (i.e. further away from zero) than the test statistic that has been observed under the assumption that the null hypothesis is true, and a p-value equal to or less than 0.05 is considered to be sufficient evidence to reject the null hypothesis. However, the analyses described in subsequent chapters consist of attempts to find associations between multiple variants and multiple phenotypes, which requires many statistical tests. The more independent statistical tests that are performed, the higher the chance of incorrectly rejecting the null hypothesis. This means that if, for example, 20 statistical tests that are not dependent on each other are performed and a p-value of 0.05 (i.e. 1 in 20) is considered sufficient evidence to reject the null hypothesis, then, by chance alone, one of those statistical tests is likely to provide sufficient evidence to reject the null hypothesis (because there are 20 tests).

To correct for this multiple testing problem, the p-value is divided by the number of independent statistical tests, which varies, so the actual correction is described within the section that it is applied. In order to determine the number of independent statistical tests, it is necessary to determine the number of independent genetic variants (if multiple genetic variants are being tested) and to determine the number of independent phenotypes. These are then multiplied to calculate the number of independent tests.

In this thesis, genetic variants are considered to be independent for the purposes of multiple testing if they are not in strong linkage disequilibrium. To determine the

number of independent variants, all pairs of variants within the locus (see Section 2.5) with $R^2 > 0.1$ were listed, one variant from each pair was removed and this process was repeated until there were no pairs of variants remaining. This was implemented using the `--indep-pairwise 60 kb 1 0.1` flag in PLINK 2.0 [66]. Although the 247 variants referred to in section 2.5 can be adequately represented by 2 variants based on the results of the analyses described in this paragraph, all 247 variants are tested regardless; the results of the `--indep-pairwise` flag are used purely for the purposes of determining the number of independent tests to correct for multiple testing.

To determine the number of independent phenotypes, a principal components analysis (PCA) is carried out. The aim of PCA is to see whether a large number of likely correlated variables can be represented by a smaller number of uncorrelated variables (referred to as principal components). Each principal component represents a certain amount of variation in the original data, with the first PC representing the highest amount of variation, the second PC representing the second highest amount of variation and so on. It is assumed that the minimum number of PCs required to represent >90% of the variation in the original data represents the number of independent phenotypes. It should be noted that the principal component analyses described here are used specifically for the purposes of determining the number of independent phenotypes and that all phenotypes are always tested regardless.

2.9 Genetic models

In this thesis, the genotype of an individual is coded using either the additive model or the dominant model or the recessive model. The additive model assumes that if carrying one copy of a minor allele produces a change in a trait, then carrying a second copy of the minor allele produces an additional change in the same trait of the same magnitude. The dominant model assumes that if carrying one copy of a minor allele produces a change in a trait, then carrying a second copy of the minor allele does not produce a further change. The recessive model assumes that carrying two copies of the minor allele produces a change in a trait whereas carrying one copy does not. The rs9536314 variant will be used to explain how the additive, dominant and recessive models are implemented. An individual can carry up to two copies of either the T (thymine) allele or the G (guanine) allele at the rs9536314 variant.

If G is the effect allele and an additive model is being used, then individuals with the TT, GT, and GG genotypes are coded as 0, 1 and 2, respectively, and these recoded genotypes are used as a predictor variable in the statistical models described earlier (see Section 2.7). If, for example, individuals who carry one copy of the G allele have a hip circumference that is 0.01 cm lower in comparison to those who carry no copies of the G allele then, under the additive model, individuals who carry two copies of the G allele would have a hip circumference that is 0.02 cm lower in comparison to those who carry no copies of the G allele and would have a hip circumference that is 0.01 cm lower in comparison to those who do carry one copy of the G allele.

If G is the effect allele and a dominant model is being used, then individuals with the TT, GT, and GG genotypes are coded as 0, 1 and 1, respectively. Under the dominant model, individuals who carry either one or two copies of the G allele would have a hip circumference that is 0.01 cm lower than those with no copies of the G allele.

If G is the effect allele and a recessive model is being used, then individuals with the TT, GT, and GG genotypes are coded as 0, 0 and 1, respectively. Under the recessive model, individuals who carry two copies of the G allele would have a hip circumference that is 0.01 cm lower in comparison to both those who carry no copies of the G allele and those who carry one copy of the G allele.

3 Klotho and Longevity

3.1 Introduction

In Section 1.2, the advantages of measuring ageing through longevity are discussed. Briefly, it is possible that individuals who live to extreme old age have unique biological traits [67] that GWASs of lifespan (briefly reviewed in section 1.4.2) may not detect. In this chapter, we use genetics to examine the evidence that Klotho is related to longevity (as opposed to lifespan). If Klotho is a valid target for anti-ageing research, then one might expect genetic variants that are located in and around the Klotho gene locus to be over- or under-represented in an elderly population in comparison to a younger control population. In this chapter, reports of previous studies that have used this approach are reviewed. This is followed by a discussion of the results of work carried out using this same approach, incorporating data from the Newcastle 85 Plus study.

Multiple studies have been carried out to explore the relationship between genetic variants at the *KL* gene locus and longevity. These have mostly been focused on a pair of functional genetic variants, in complete linkage disequilibrium, that result in F352V (rs9536314) and C370S (rs9527025) substitutions, first reported by Arking *et al*. [68], and referred to as the *KL*-VS haplotype.

Arking *et al*. [68] went on to report that *KL*-VS heterozygotes were more common in Bohemian Czechs aged ≥75 years than in newborn controls, and also found that *KL*-VS heterozygotes became more common with age in Ashkenazi Jews aged ≥79 years [69]. Invidia *et al*. [70] also reported that *KL*-VS heterozygotes were more common in elderly Italian individuals (mean age 78 years) in comparison to younger controls (mean age 53 years).

However, other studies were not able to replicate the aforementioned longevity advantage reported in *KL*-VS heterozygotes. Arking *et al*. [68] were not able to replicate their findings from Bohemian Czechs in either Baltimore-based Caucasians or in Baltimore-based African-Americans, though it should be noted that participants defined as long-lived were only ≥65 years old as opposed to ≥75 years old. Novelli *et al*. [71]

compared US participants aged between 99 and 111 years old to controls aged <35 years and Flachsbart *et al.* [72] compared German centenarians to younger controls (60-75 years old), but neither group found evidence of a difference in allele and/or genotype frequencies between long-lived cases and younger controls.

These conflicting reports indicate that, at a population level, the relationship between the *KL*-VS haplotype and longevity remains unclear. We aim to explore this area further by comparing long-lived individuals from the Newcastle 85 Plus study with middle-aged Newcastle-based controls from the UK Biobank to test whether or not variants at the *KL* gene locus are over- or under-represented in older individuals.

3.2 The Newcastle 85 Plus study

We obtained data from long-lived cases from the Newcastle 85 Plus (N85+) study which, in 2006, recruited 1042 participants born in 1921 regardless of their health status (excepting those with late-stage terminal illness), including those with cognitive impairment (for whom careful procedures were followed to secure proxy consent, where appropriate). All individuals who met the recruitment criteria and were not suffering from a terminal illness were eligible. The N85+ study was approved by the Newcastle and North Tyneside 1 research ethics committee (reference number 06/Q0905/2) [73].

N85+ participants had been genotyped using Illumina Omni genotyping arrays. The details of the QC carried out on the genetic data from the N85+ study are available in Deelen *et al.* [74]. In addition to this QC, all variants with an INFO score of <0.8 were excluded for the work described in this study.

3.3 Statistical Analyses

We used R 4.0.2 [75] to carry out analyses, unless stated otherwise. We used qctool [76] and GTOOL [77] to convert both the UKB and the N85+ imputed data to hardcalls, using a posterior probability threshold of 0.9.

We used the chi-squared (χ^2) test to compare genotype counts between Newcastle UKB participants (i.e. UKB participants who attended the Newcastle assessment centre aged

≥45 years and <65 years) and N85+ participants. We used the Z test of proportions, as implemented in the prop.test(...,correct=FALSE) function in R, to compare the proportion of rs9536314 carriers and rs9536314 heterozygotes between the two aforementioned groups. For rs9536314, we tested four different genotype models (additive, dominant, recessive and heterozygous), so the multiple-testing corrected p-value threshold we used was 0.05/4 = 0.0125.

If the genotype distribution at any variant (except rs9536314, where all genotype models were tested regardless) was found to differ significantly between the N85+ cohort and the Newcastle UKB cohort, we identified the underlying genetic model and then used 2×2 contingency tables to compare the proportion of N85+ participants (the "event") between the two genotype/allele groups (the "exposed" and "unexposed" groups). Odds ratios and the corresponding 95% confidence intervals and p-values were generated as described by Szumilas [78].

To replicate any positive results, we separated the non-Newcastle UKB participants by the assessment centre location that they attended at baseline (UKB field 54) to account for population stratification, compared the proportion of UKB participants aged ≥80 years (as of April 26, 2020) between two genotype/allele groups and meta-analysed the results using the Mantel-Haenszel method as implemented in the metabin(...,sm="OR",method="MH") function from the "meta" package in R [79].

3.4 Results

After QC, there were 642 N85+ participants (60.6% female) and 18,295 Newcastle UKB participants (54.9% female) remaining. There are no statistically significant associations between *KL* variants and sex across the UKB (Supplementary Table 3.1).

Since the two variants making up the *KL*-VS haplotype are well-characterised functional *KL* variants in humans, we investigated whether or not their genotype distributions differed significantly between N85+ participants and Newcastle UKB participants (Table 3.1). We found no significant difference in the genotype distribution for rs9536314 between the two cohorts (TT, TG, GG: [69.5%, 27.6%, 3.0%] vs. [70.4%, 27.1%, 2.5%], p

= 0.74). It should be noted that only the results for rs9536314 are provided because rs9536314 and rs9527025 are in complete linkage disequilibrium ($R^2 = 1$).

Table 3.1 Allele/genotype distribution at rs9536314 in N85+ and UKB

Distribution of alleles and genotypes at rs9536314 Newcastle 85 Plus (N85+) participants and middle-aged Newcastle-based UK Biobank (N_UKB) participants.

	N85+		N_UKB	
Genotype				
TT	446	69.5%	12880	70.4%
TG	177	27.6%	4954	27.1%
GG	19	3.0%	461	2.5%
Additive				
T	1069	83.3%	30714	83.9%
G	215	16.7%	5876	16.1%
Dominant				
TT	446	69.5%	12880	70.4%
TG/GG	196	30.5%	5415	29.6%
Recessive				
TT/TG	623	97.0%	17834	97.5%
GG	19	3.0%	461	2.5%
Heterozygous				
TT/GG	465	72.4%	13341	72.9%
TG	177	27.6%	4954	27.1%

It has been suggested that *KL*-VS heterozygotes are at an advantage when it comes to longevity. We therefore compared the proportion of rs9536314 heterozygotes between N85+ participants and Newcastle UKB participants, but found no difference (27.6% vs 27.1%, p = 0.79). We also compared the proportion of rs9536314 GG homozygotes and rs9536314 G carriers between the two cohorts, but again did not find any differences (3.0% vs. 2.5%, p = 0.49; and 30.5% vs. 29.6%, p = 0.61). The allele frequencies were also similar (16.7% vs 16.1%, p = 0.51). The results are summarised in Table 3.1.

We next sought to compare the genotype distributions between the N85+ participants and the Newcastle UKB participants for the remaining 194 *KL* variants (Supplementary Table 3.2). The genotype distributions of rs2283368 and rs9536338 differ between the two cohorts ($p = 2.1 \times 10^{-3}$ and $p = 7.5 \times 10^{-3}$, respectively). These variants were selected for further analysis.

We found that individuals from the N85+ study were more likely to be present in the rarer rs2283368 CC group than the rs2283368 TT/TC group (OR = 2.42, [95% CIs 1.44 to 4.06, $p = 4.0 \times 10^{-4}$]), which suggests that the CC genotype could be associated with longevity (Table 3.2). However, when we attempted to replicate this result by comparing the proportion of UKB participants aged ≥80 years who were present in the rs2283368 CC group to the proportion who were present in rs2283368 TT/TC group at each assessment centre across the UK (Figure 3.1), we found no statistically significant difference (random effects model: OR = 1.15 [95% CIs 0.96 to 1.37, p = 0.14]).

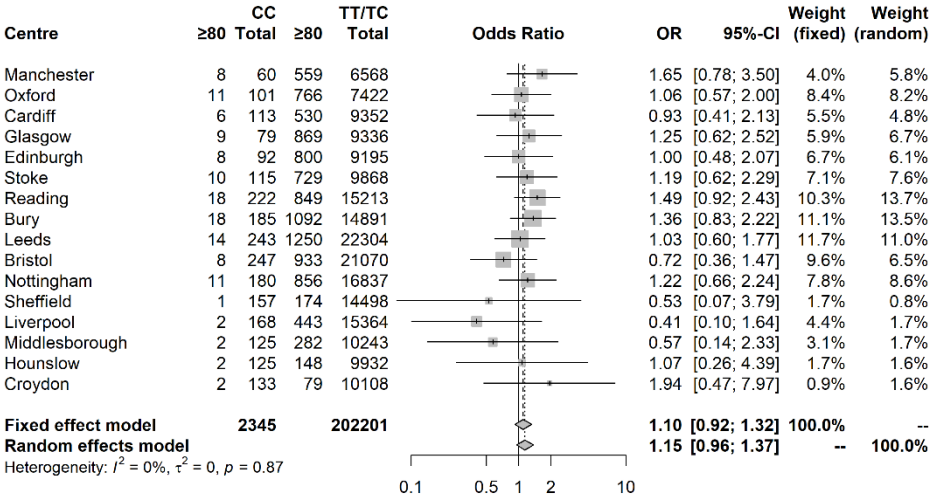
Table 3.2 Distribution of rs2283368 CC genotypes in N85+ and UKB

Distribution of rs2283368 CC genotypes amongst Newcastle 85 Plus (N85+) participants and middle-aged Newcastle-based UK Biobank (N_UKB) participants.

	CC	TT/TC
N85+	16	626
N_UKB	184	17435
	200	18061

Figure 3.1 Meta-analysis of rs2283368 CC genotype distribution in the UKB

Meta-analysis of the distribution of rs2283368 CC genotypes amongst long-lived cases and younger controls in the UKB.



We also found that the rarer rs9536338 G allele is less frequent in N85+ participants (OR = 0.81 [95% CIs 0.72 to 0.92, p = 6.3×10^{-3}]), which suggests that the G allele may be associated with reduced longevity (Table 3.3). We attempted to replicate this finding by comparing the proportion of UKB participants aged ≥80 years who were present in the rs9536338 G group to the proportion who were present in the rs9536338 C group at each assessment centre (Figure 3.2), but we did not find any evidence to support our initial result (random effects model: OR = 1.02 [95% CIs 0.99 to 1.05, p = 0.2]).

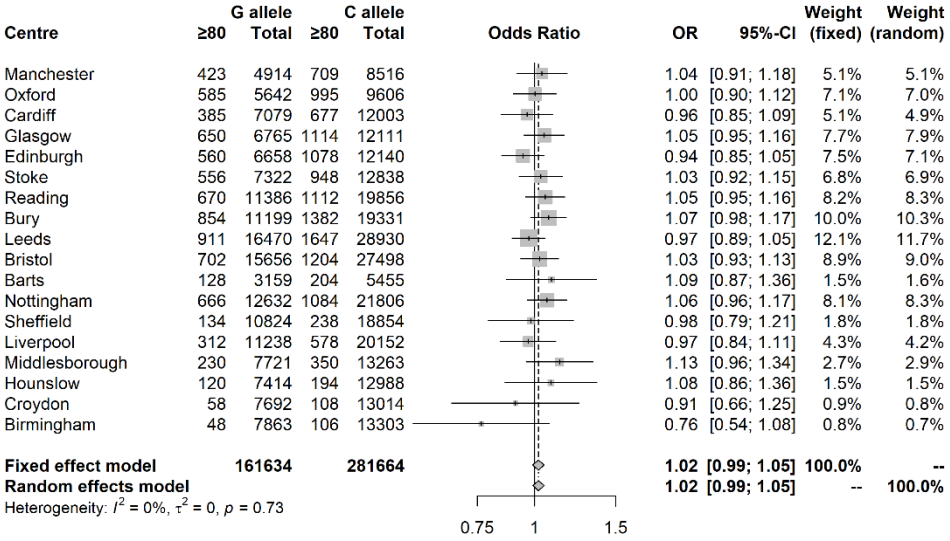
Table 3.3 Distribution of G and C alleles at rs9536338 in N85+ and UKB

Distribution of G and C alleles at rs9536338 amongst Newcastle 85 Plus (N85+) participants and middle-aged Newcastle-based UK Biobank (N_UKB) participants.

	G	C
N85+	352	768
N_UKB	12785	22697
	13137	23465

Figure 3.2 Meta-analysis of G/C allele distribution at rs9536338 in the UKB

Meta-analysis of the distribution of G and C alleles at rs9536338 amongst long-lived cases and younger controls in the UKB.



3.5 Discussion

In this chapter, we sought to verify previous reports of associations between the *KL*-VS haplotype and longevity and to identify novel variants at the Klotho gene locus that may also be associated with longevity, if any. Our data do not support the presence of an association between rs9536314, a genetic variant that characterises the *KL*-VS haplotype, and longevity. Although we identified possible associations with the rs2283368 and rs9536338 variants, we were unable to replicate these in a second, much larger, sample.

Arking *et al.* [68] reported an association between rs9536314 and longevity, but we could not identify this in our data. A possible reason for this may be that Arking *et al.* [68] compared newborns to elderly participants, which means that the effect they observed could be explained by a relationship between rs9536314 and infant mortality as opposed to longevity. Indeed, two other studies [71,72], in which adults, rather than newborns, were used as controls, also did not provide evidence for an association between longevity and rs9536314.

We were unable to replicate the associations that we found between rs2283368 and rs9536338 and longevity, so they are likely to be false positives. However, it is also possible that the lack of UKB participants aged ≥85 years and the consequent need to re-define long-lived cases as those aged ≥80 years may have reduced the power of our replication sample to detect an effect, if present.

Previous studies have used a variety of age thresholds to define their long-lived cases and their younger controls, which makes it difficult to compare them and to establish a pattern. We chose to use 85 years as the threshold to define our long-lived cases and this seems reasonable given that the pre-pandemic life expectancy in most countries, including the UK [80], has not yet exceeded 85 years and previous publications frequently consider those aged above 85 years as the oldest old [81]. We defined our middle-aged controls as those aged between 45 and 64 years inclusive because this is the current MeSH definition of middle-aged [82]. Another approach, used by Invidia *et al.* [83], involves generating population-specific survival curves and defining thresholds based upon the ages at which mortality increases or decreases. However, these ages

are likely to vary between populations (and therefore will be subject to population-specific biases) and will be affected by events such as the COVID-19 pandemic [84].

There are some potential limitations in our work. Participants in the UKB study are reportedly healthier [85] than the average for a person from the UK population. It could be argued that individuals who are likely to be long-lived tend to be free of any age-related morbidities until the very end of their life [86], so the UKB sample may contain a higher proportion of individuals who will ultimately be long-lived. The Newcastle UKB cohort, which was the control sample in this study, may therefore contain some individuals who, in time, would be included as cases in studies such as ours. Furthermore, the N85+ Study sought to recruit participants regardless of health status (excepting those with late-stage terminal illness), including those with cognitive impairment (for whom careful procedures were followed to secure proxy consent, where appropriate), which reduces the difference between the long-lived cases and younger controls because the maintenance of cognitive independence into very old age is a characteristic of longevity. Together, these factors may have reduced the power of our discovery sample to detect an effect, if present.

In conclusion, we did not find sufficient evidence to support the previously reported associations between *KL*-VS and longevity. Once further follow-up data from the UKB becomes available as the cohort gets older and some individuals begin to exceed the average lifespan, the associations between rs2283368 and rs9536338 and longevity should be re-tested. However, despite the novel (albeit unreproduced), associations that we describe, the evidence, at least on a population genetics level, remains fragmented. Thus, the results described in this chapter do not reliably support the role of *KL* as a longevity factor.

3.6 Postface

As of 13th August 2021, the work described in this chapter is under consideration for publication by *The Journals of Gerontology*, *Series A*.

4 Klotho and Cardiometabolic Traits

4.1 Introduction

As discussed in Chapter 3, there does not appear to be any reliable evidence of an association between Klotho variants and longevity at age 85 years or over. However, it is also important to explore the possibility that Klotho variants may be associated with specific diseases that are amongst the most common causes of mortality amongst older people.

Cardiometabolic diseases are a group of pathologies that include cardiovascular disease and the metabolic syndrome. The metabolic syndrome is characterised by central adiposity (as measured by waist circumference), dyslipidaemia (namely increased triglycerides and reduced high-density lipoprotein), hypertension and insulin resistance (which progresses into type 2 diabetes) [87]. Although the exact diagnostic criteria for metabolic syndrome are debated [87], the characteristics of metabolic syndrome are associated with ischaemic heart disease and stroke [88], which, in 2018, were the second and fourth leading causes of death in UK residents aged 65 years or more, respectively [89].

The chapter begins with a review of the previous evidence for an association between Klotho variants and cardiometabolic diseases and traits. Then, an attempt is made to verify the reported associations between *KL* variants, including the *KL*-VS haplotype, and cardiometabolic phenotypes in UKB participants, by carrying out a phenome scan of cardiometabolic outcomes and their associated risk factors and to search for novel associations between the *KL* genetic variants and cardiometabolic phenotypes using the same approach. The results of these analyses are also presented and discussed in this chapter.

4.2 Klotho variants and cardiometabolic phenotypes

There is conflicting evidence that genetic variants at the Klotho gene locus may be associated with cardiometabolic diseases and their risk factors. The ones that are most

commonly assessed are: *KL*-VS, a haplotype represented by two variants in complete linkage disequilibrium that result in F352V (rs9536314) and C370S (rs9527025) substitutions; G395A (rs1207568); and C1818T (rs564481).

4.2.1 *KL*-VS

Since rs9536314 and rs9527025, represented by the *KL*-VS haplotype, are the most well-characterised Klotho variants, their association with cardiovascular outcomes and risk factors is reviewed. *KL*-VS was first reported to be associated with coronary artery disease (CAD) by Arking *et al*. [90], who found that Caucasian carriers of the rs9536314 G allele were at a lower risk of early-onset CAD (n = 520), but Low *et al*. [91] did not find any such association (n = 440). Paula *et al*. [56] found that *KL*-VS is not associated with CAD in an older (≥60 years old) Brazilian population (n = 168) but Donate-Correa *et al*. [58] found a higher rate of atherosclerosis in F352V carriers in comparison to non-carriers in a Spanish population (n = 105). Arking *et al*. [69] found that *KL*-VS heterozygotes were less likely to experience a stroke, but: Majumdar *et al*. [92] found the opposite to be true in a South Asian population (n = 1034); and Paula *et al*. [56] did not find an association between stroke and *KL*-VS in their data. *KL*-VS also does not appear to be associated with myocardial infarction [69,93]. With respect to diabetes, the V allele is reportedly associated with a lower risk of diabetes in Brazilians [56], but not in Caucasians [94]. The available evidence for an association between *KL*-VS cardiometabolic outcomes is contradictory.

In terms of cardiovascular risk factors, Arking *et al*. [69] found that *KL*-VS heterozygotes have a lower systolic blood pressure (SBP) in their data but, in their respective datasets: Nzietchueng *et al*. [95] found this effect to be restricted to VV homozygotes, Donate-Correa *et al*. [58] did not find evidence for an association between blood pressure and *KL*-VS in their data; and Majumdar and Christopher [96] actually found that VV homozygotes have a high SBP in a South Asian population. Paroni *et al*. [97] found that VV homozygotes had higher levels of high-density lipoprotein (HDL), but three other groups did not find evidence of this association in their respective data [56,69,94]. The evidence for an association between cardiovascular risk factors, namely SBP and HDL, and *KL*-VS is contradictory.

4.2.2 G395A (rs1207568)

There is evidence that G395A may affect the risk of CAD. Imamura *et al*. [98], found that the 395A allele is associated with CAD in a Japanese population. These findings were corroborated in a Korean population by Jo *et al*. [99], who found that carriers of the 395A allele are more likely to suffer from CAD. In addition to its effects on CAD risk, the 395A allele has been associated with an increased risk of cardioembolism in Korean women [100], but Pereira *et al*. [93] did not find evidence of a relationship between the G395A variant and stroke in a Brazilian population.

There is evidence that the effect of G395A on CAD risk could be age-dependent. Rhee *et al*. [101] found no evidence of an association between the 395A allele and coronary artery stenosis (mean age = 58.5 years) and in fact observed that, when the analysis was restricted to those aged over 60 years, stenosis of the coronary artery is actually less common in Koreans who carry the 395A allele. In contrast, when Jo *et al*. [99] restricted their analyses to Koreans aged over 60 years, there was no association between 395A and CAD and in those aged under 60, CAD was more common in 395A carriers. Together, this suggests that the effect of the 395A allele on CAD could be age-dependent, at least in Koreans, but the possibility of this being a false positive cannot be discounted.

Rhee *et al*. [102] found that 395A carriers had a higher SBP in comparison to non-carriers in a sample of Korean women and the absence of a statistically significant effect in a mixed sex sample of Koreans [101] raises the possibility that this effect is female-specific, at least in Koreans. However, the G395A variant does not appear to be associated with hypertension in a Spanish population [58] and Wang *et al*. [57] found that the 395A allele lowers the odds of hypertension in a Chinese Han population. The latter was corroborated by Gao *et al*. [103], who used data from elderly Chinese participants and found that carriers of the 395A allele had a lower SBP. Shimoyama *et al*. [104] found that, in a Japanese population, male and female carriers of the 395A allele had lower HDL and higher fasting glucose levels, but Paroni *et al*. [97] did not find evidence for either of these associations in an Italian population. Donate-Correa *et al*. [58] found that carriers of the 395A were more likely to get diabetes, but Gao *et al*. [103] did not find evidence of such an association. In summary, the effect of the G395A

variant on cardiometabolic risk factors varies with ethnicity and perhaps with sex but, once again, the possibility of this being a false positive cannot be discounted.

4.2.3 C1818T (rs564481)

The third of the three commonly assessed variants is C1818T (rs564481) and, as is the case with *KL*-VS and G395A, its effects also vary between different ethnicities, though this variation could again be a case of false positives. Rhee *et al.* [102] and Shimoyama *et al.* [104] respectively found that Korean women and Japanese women who carried the T allele have higher fasting glucose levels but interestingly, Paroni *et al.* [97] found that the T allele had the opposite effect on fasting glucose in Italian women. The association between C1818T and fasting glucose is not present in South Asians [92,96]. Rhee *et al.* [101] found that Koreans who carry the T allele had a reduced risk of CAD, but Pereira *et al.* [93] found that Brazilians who carry the T allele are more likely to get MI. C1818T is not associated with stroke in Korean, South Asian and Brazilian populations [92,93,100].

4.2.4 Summary

To summarise: the associations between *KL*-VS, G395A and C1818T and cardiometabolic phenotypes could vary between ethnic groups; the associations between G395A and C1818T and cardiometabolic phenotypes may vary with age and sex, respectively; however, since the evidence for these genetic variants having different effects across ethnic groups, age and sex is from small sample sizes, one cannot discount the possibility of these patterns being false positives. As part of the work described in this chapter, an attempt is made to verify these reported associations by carrying out a phenome scan of cardiometabolic outcomes and their associated risk factors in a much larger sample available from the UKB.

4.3 Phenotypes

In order to assess whether Klotho variants are associated with cardiovascular outcomes and their risk factors, it was necessary to use both measurements performed during the assessment centre as well as the summarised data generated from the participants' hospital records (see Section 2.2) provided by UKB to generate phenotypes.

4.3.1 Diabetes and glycaemic traits

Type 2 diabetes: participants who reported being diagnosed with diabetes by a doctor (UKB field 2443, which records participants' responses when asked to declare whether they had been diagnosed with diabetes by a doctor) were considered as cases; and participants with one or more primary diagnoses of ICD-10 codes E10 (Insulin-dependent diabetes mellitus), E12 (Malnutrition-related diabetes mellitus), E13 (Other specified diabetes mellitus), E14 (Unspecified diabetes mellitus) and O24 (Diabetes mellitus in pregnancy, childbirth, and the puerperium) listed in UKB field 41202 (which lists all lifetime hospital diagnoses for each participant since records were available) were set to missing to ensure that the analyses were restricted to type 2 diabetes.

Type 2 diabetes can be diagnosed on the basis of a fasting glucose test. However, a fasting glucose test does not necessarily provide an accurate representation of longer-term blood glucose levels. Chronically elevated glucose levels result in the production of advanced glycation end-products [105], which is where the excess glucose binds to various proteins and lipids. One protein that glucose can bind to is haemoglobin, which produces glycated haemoglobin (HbA1c). The ratio of non-glycated haemoglobin to glycated haemoglobin is used a diagnostic marker for type 2 diabetes since a higher concentration of HbA1c indicates chronic exposure to a higher level of glucose [106].

Standard clinical biochemistry analyses were performed using blood samples provided by participants at the UKB assessment centres. Fasting glucose concentrations for participants were obtained from UKB field 30740 and participants who reported fasting for less than 8 hours (UKB field 74, which asked participants how many hours had passed since they had last eaten) were excluded. Glycated haemoglobin (HbA1c) concentrations were obtained from UKB field 30750. Since participants who have been diagnosed with type 2 diabetes are likely to have been prescribed anti-diabetic medication, only participants who were specified as controls for the type 2 diabetes phenotype were used for analyses of fasting glucose and HbA1c phenotypes in order to minimise the effects of any medical interventions against diabetes.

4.3.2 Incident cardiovascular and stroke outcomes

As described in Section 2.2, UKB provides a list of diseases that each participant was diagnosed with and the date on which they were diagnosed with each disease. A participant with one or more diagnoses of a particular disease was considered to be a case for that disease and all other participants were considered to be controls. Some diseases are similar to the disease of interest but have a different ICD-10 code (e.g. atherosclerotic heart disease, I25.1, is a type of ischaemic heart disease, I20-I25), so a participant may be considered as a control for the disease of interest yet may also have been diagnosed with a similar disease, which means that this participant's case-control status is arguably ambiguous. These participants are therefore set to missing.

In this chapter, only incident cases of cardiovascular and stroke outcomes are considered. The purpose of this is to try and reduce the bias that exists because of the fact that people who get premature CAD would be less likely to volunteer for the UKB study, partly because some of those who suffer from premature CAD may have died.

Table 4.1 shows ICD-10 codes that were used to define cardiovascular and stroke outcomes and to set controls to missing. To restrict the analyses to incident cases, data from UKB fields 41202 (Diagnoses - main ICD10) and 41262 (Date of first in-patient diagnosis - main ICD10) were used to exclude participants with one or more primary diagnoses of diseases affecting the circulatory system (ICD-10 codes I00 to I99, G45 and G46) before attending the assessment centre at baseline (UKB field 53). To generate the premature CAD phenotype, participants who were considered as cases for the CAD phenotype were set as controls if their first primary diagnosis of ICD-10 code I251 occurred on or after 60 years of age; this age cutoff was chosen because NICE defines premature CAD as 'an event before 60 years' [107].

Table 4.1 ICD-10 codes used to define cardiovascular outcomes

ICD-10 codes from UKB field 41202 used to define cases and to exclude controls from the analyses.

Phenotype	Cases	Controls set to missing
Coronary artery disease (CAD)	I251 Atherosclerotic heart disease	I20-I25 Ischaemic heart diseases
Acute myocardial infarction (MI)	I21 Acute myocardial infarction	I20-I25 Ischaemic heart diseases
Stroke	I60 Subarachnoid haemorrhage I61 Intracerebral haemorrhage I62 Other nontraumatic intracranial haemorrhage I63 Cerebral infarction I64 Stroke, not specified as haemorrhage or infarction G45 Transient cerebral ischaemic attacks and related syndromes G46 Vascular syndromes of brain in cerebrovascular diseases	I60-I69 Cerebrovascular diseases
Cerebrovascular disease	I60-I69 Cerebrovascular diseases G45 Transient cerebral ischaemic attacks and related syndromes G46 Vascular syndromes of brain in cerebrovascular diseases	

4.3.3 Cardiovascular risk factors

Systolic blood pressure (SBP) and diastolic blood pressure (DBP) measurements taken during the assessment centre visit at baseline were obtained from UKB fields 4080 and 4079, respectively. 15 mmHg and 10 mmHg was added to the SBP and DBP measurements for participants who reported taking antihypertensive medication (UKB fields 6153 and 6177) because this has been shown to increase statistical power [108,109]. Waist circumference measurements and body mass index (BMI), also taken during the baseline assessment centre visit, were obtained from UKB fields 48 and 21001, respectively. Data from the clinical biochemistry analyses also included measurements for direct LDL (low-density lipoprotein cholesterol), HDL (high-density lipoprotein cholesterol), triglycerides, total cholesterol and LpA (lipoprotein A) were obtained from UKB fields 30780, 30760, 30870, 30690 and 30790, respectively.

4.4 Statistical analyses

PLINK 2.0 [66] was used to fit additive linear models between the cardiometabolic phenotypes and the genetic variants in all individuals. For *KL*-VS, G395A and C1818T, dominant and recessive linear models were also fitted because previous studies revealed evidence of non-additive effects. Unless otherwise specified, all association analyses were adjusted for the first four genetic principal components (PCs) (UKB field 22009) and the genotyping array that the participant was genotyped on. We used the first four genetic PCs because they represent >90% variation in the genome. Any quantitative phenotypes and covariates were standardised to a mean of 0 and a variance of 1 before any linear modelling was performed.

To find conditionally independent associations, we used a forwards-backwards stepwise selection procedure as implemented in the ols_step_both_p() function from the olsrr package in R [110]. The procedure is applied only to variants that are found to be individually associated with a phenotype and is as follows: one, the variant with lowest p-value is added to the model; two, if any variants are statistically insignificant ($p > 0.05$), the are removed from the model; and three, steps one and two are repeated until none of the variants that were originally associated with the phenotype remain.

To test whether the effect of *KL*-VS, G395A and C1818T on the cardiometabolic phenotypes varied with age, PLINK 2.0 was used to run additive linear models in those aged ≥60 years at baseline and in those aged <60 years at baseline (this was not done for the premature CAD phenotype since there are no cases in those aged ≥60 years at baseline based upon the definition of premature CAD provided in the Phenotypes section). For each phenotype-variant combination, we used the metagen() function from the meta package in R [79] to perform a chi-squared test of heterogeneity between the effect sizes from the two strata. The aforementioned procedure was also used to test whether there is heterogeneity in the effect of *KL*-VS, G395A and C1818T on cardiometabolic phenotypes between males and females.

We applied statistical correction for multiple testing since we tested for associations between multiple variants and multiple phenotypes (see Section 2.8). In order to determine the number of independent phenotypes, a principal component (PC) analysis was performed, which showed that 8 PCs are needed to represent >90% of the variation in the 11 quantitative cardiometabolic phenotypes. The 8 PCs referred to here were generated from the 11 quantitative cardiometabolic phenotypes as opposed to the genetic PCs referred to in the previous paragraph, which were generated by UKB from the genotype data. PCA cannot be applied to the binary cardiometabolic phenotypes, so they were assumed to be independent for the purposes of multiple testing. Two independent variants remain after applying the procedure described in Section 2.8. *KL*-VS, G395A and C1818T are treated as 3 independent variants for the purposes of multiple testing when analysed separately from the other 211 variants. A *p*-value threshold of 0.05 was used and Bonferroni-correction for the appropriate number of independent tests in each case was applied when necessary.

4.5 Results

4.5.1 Population characteristics

There are 335,024 participants that pass the QC described in Chapter 2. The sample has a mean age of 57 ± 8 years at baseline and 53.6% are female. Table 4.2 and Table 4.3 summarise the binary phenotypes (disease traits) and the quantitative phenotypes in the sample, respectively.

Table 4.2 Cardiometabolic disease traits

Number of incident cases and controls for binary cardiometabolic phenotypes. CAD = coronary artery disease.

Phenotype	Controls	Cases
Premature CAD	296907	1035
CAD	292365	5577
Myocardial infarction	292365	3048
Stroke	297603	2825
Cerebrovascular disease	297603	4045
Type 2 diabetes	318023	15449

Table 4.3 Quantitative cardiometabolic traits

Distribution of quantitative cardiometabolic phenotypes in their original units. N = sample size. SD = standard deviation.

Phenotype	N	Mean	SD
SBP	305630	141.3	20.6
DBP	305637	84.3	11.2
HbA1c	303003	35.1	4.4
Fasting glucose	10619	5.1	0.7
Triglycerides	319183	1.8	1
LDL	318837	3.6	0.9
HDL	292497	1.5	0.4
Total cholesterol	319428	5.7	1.1
Lipoprotein A	253988	44.1	49.5
Waist circumference	334475	90.4	13.5
BMI	333960	27.4	4.8

4.5.2 Association of cardiometabolic outcomes with *KL*-VS, G395A and C1818T

Since *KL*-VS (rs9536314), G395A (rs1207568) and C1818T (rs564481) are the Klotho variants that are most often reported to be associated with cardiometabolic traits (see Section 4.2), initial analyses were carried out to see whether any of these three variants are associated with cardiovascular outcomes or type 2 diabetes under additive, dominant and recessive models, adjusted for age and sex. A *p*-value threshold of 0.05/54 was used because the effects of three variants on six outcomes under three different models (3 × 6 × 3 = 54 independent tests) were tested. None of the three variants are associated at a *p*-value of 0.05/54 with any of the six outcomes regardless of which model was used (Table 4.4). The results for the unadjusted additive model are provided in Supplementary Table 4.1.

Table 4.4 Regression of cardiometabolic outcomes on *KL* variants

Log odds ratios (beta) and their respective errors (SE) when regressing the six cardiometabolic outcomes on *KL*-VS (rs9536314), G395A (rs1207568) and C1818T (rs564481) using logistic regression. N = sample size. A1 = effect allele.

Trait	Variant	A1	N	Additive			Dominant			Recessive		
				beta	SE	*p*	beta	SE	*p*	beta	SE	*p*
CAD	rs1207568	A	297942	-0.0023	0.025	0.93	0.01	0.029	0.73	-0.084	0.075	0.26
CAD	rs9536314	G	297942	-0.014	0.026	0.59	-0.011	0.03	0.72	-0.068	0.089	0.44
CAD	rs564481	T	297942	-0.018	0.02	0.36	0.00042	0.029	0.99	-0.067	0.038	0.078
Cerebrovascular disease	rs1207568	A	301648	0.014	0.029	0.63	0.0069	0.034	0.84	0.073	0.082	0.37
Cerebrovascular disease	rs9536314	G	301648	-0.032	0.031	0.31	-0.047	0.035	0.18	0.053	0.098	0.59
Cerebrovascular disease	rs564481	T	301648	0.0032	0.023	0.89	-0.0068	0.033	0.84	0.023	0.043	0.6
Myocardial infarction	rs1207568	A	295413	-0.017	0.033	0.62	-0.021	0.039	0.59	-0.012	0.098	0.9
Myocardial infarction	rs9536314	G	295413	0.026	0.035	0.46	0.029	0.04	0.47	0.037	0.11	0.75
Myocardial infarction	rs564481	T	295413	-0.04	0.027	0.13	-0.048	0.038	0.21	-0.063	0.051	0.21
Premature CAD	rs1207568	A	297942	0.052	0.056	0.35	0.063	0.066	0.34	0.059	0.16	0.71
Premature CAD	rs9536314	G	297942	-0.053	0.061	0.38	-0.036	0.069	0.6	-0.3	0.23	0.18
Premature CAD	rs564481	T	297942	-0.045	0.045	0.32	-0.074	0.065	0.26	-0.033	0.086	0.7
Stroke	rs1207568	A	300428	0.028	0.034	0.4	0.031	0.04	0.44	0.052	0.098	0.6
Stroke	rs9536314	G	300428	-0.0041	0.037	0.91	-0.017	0.042	0.69	0.092	0.12	0.43
Stroke	rs564481	T	300428	-0.021	0.028	0.44	-0.035	0.04	0.37	-0.016	0.052	0.76
Type 2 diabetes	rs1207568	A	333472	-0.023	0.015	0.13	-0.014	0.018	0.43	-0.12	0.046	0.011
Type 2 diabetes	rs9536314	G	333472	-0.0074	0.016	0.65	0.0028	0.018	0.88	-0.11	0.055	0.051
Type 2 diabetes	rs564481	T	333472	-0.016	0.012	0.18	-0.011	0.017	0.54	-0.04	0.023	0.079

4.5.3 Association of cardiometabolic traits with *KL*-VS, G395A and C1818T

We next tested to see whether *KL*-VS (rs9536314), G395A (rs1207568) and C1818T (rs564481) are associated with quantitative cardiometabolic risk factors (Table 4.5). Although we tested 11 traits, we determined that >90% of the variation is explained by eight principal components (see Section 4.4), so we used a *p*-value threshold of 0.05/72 (3 × 8 × 3 = 72 independent tests).

We found that the rs9536314 G allele is associated with increased levels of HbA1c under both the additive model (beta = 0.012 [SE = 0.0034, $p = 5.7 \times 10^{-4}$]) and the dominant model (beta = 0.014 [SE = 0.0039, $p = 3.7 \times 10^{-4}$]).

We sought independent support for our HbA1c association results using publicly available data. However, there is no evidence for this association (beta = 0.0023 [SE = 0.0029, $p = 0.43$]) in the largest external GWAS of HbA1c in Europeans [111]. The results for the unadjusted additive model are provided in Supplementary Table 4.2.

Table 4.5 Regression of cardiometabolic traits on *KL* variants

Regression coefficients (beta) and their respective errors (SE) when regressing the 11 cardiometabolic traits on *KL*-VS (rs9536314), G395A (rs1207568) and C1818T (rs564481) using linear regression. N = sample size. A1 = effect allele.

Trait	Variant	A1	N	Additive			Dominant			Recessive		
				beta	SE	*p*	beta	SE	*p*	beta	SE	*p*
BMI	rs1207568	A	333960	0.0021	0.0031	0.49	0.0012	0.0037	0.74	0.011	0.0091	0.23
BMI	rs9536314	G	333960	-0.0067	0.0033	0.044	-0.0063	0.0038	0.096	-0.02	0.011	0.068
BMI	rs564481	T	333960	0.0048	0.0025	0.055	0.0027	0.0036	0.45	0.012	0.0047	0.0082
Diastolic BP	rs1207568	A	305637	0.0038	0.0032	0.24	0.0034	0.0038	0.37	0.011	0.0094	0.22
Diastolic BP	rs9536314	G	305637	-0.00085	0.0034	0.8	-0.00067	0.0039	0.86	-0.0036	0.011	0.75
Diastolic BP	rs564481	T	305637	-0.002	0.0026	0.42	-0.006	0.0037	0.11	0.0029	0.0048	0.55
Fasting glucose	rs1207568	A	10619	0.0063	0.017	0.71	0.0046	0.02	0.82	0.026	0.05	0.61
Fasting glucose	rs9536314	G	10619	-0.0035	0.019	0.85	0.00081	0.021	0.97	-0.046	0.061	0.46
Fasting glucose	rs564481	T	10619	0.0051	0.014	0.71	0.0062	0.02	0.76	0.0077	0.026	0.77
HbA1c	rs1207568	A	303003	-0.01	0.0032	0.0014	-0.012	0.0038	0.0015	-0.014	0.0093	0.13
HbA1c	rs9536314	G	303003	0.012	0.0034	0.00057	0.014	0.0039	0.00037	0.011	0.011	0.31
HbA1c	rs564481	T	303003	-0.0048	0.0025	0.06	-0.0064	0.0037	0.083	-0.0062	0.0048	0.2
HDL	rs1207568	A	292497	0.0048	0.003	0.12	0.0055	0.0036	0.13	0.0073	0.0089	0.42
HDL	rs9536314	G	292497	-0.0019	0.0033	0.56	-0.00095	0.0037	0.8	-0.012	0.011	0.25
HDL	rs564481	T	292497	0.0053	0.0024	0.031	0.0058	0.0035	0.1	0.009	0.0046	0.05
LDL	rs1207568	A	318837	0.0049	0.0032	0.12	0.0022	0.0038	0.57	0.029	0.0094	0.0018
LDL	rs9536314	G	318837	0.0057	0.0034	0.093	0.0051	0.0039	0.19	0.019	0.011	0.088
LDL	rs564481	T	318837	0.00061	0.0026	0.81	-0.0025	0.0037	0.5	0.0064	0.0048	0.18
Lipoprotein A	rs1207568	A	253988	-0.00014	0.0036	0.97	7.10E-06	0.0042	1	-0.0013	0.011	0.9
Lipoprotein A	rs9536314	G	253988	0.0013	0.0038	0.74	7.00E-04	0.0044	0.87	0.0081	0.013	0.52
Lipoprotein A	rs564481	T	253988	-0.0042	0.0029	0.15	-0.0076	0.0042	0.069	-0.0021	0.0054	0.7
Systolic BP	rs1207568	A	305630	0.0024	0.003	0.43	0.0026	0.0035	0.47	0.0045	0.0088	0.61
Systolic BP	rs9536314	G	305630	-0.0043	0.0032	0.18	-0.005	0.0036	0.17	-0.0043	0.011	0.69
Systolic BP	rs564481	T	305630	0.0018	0.0024	0.45	0.0024	0.0035	0.5	0.0025	0.0045	0.59
Total cholesterol	rs1207568	A	319428	0.0066	0.0031	0.037	0.0046	0.0037	0.22	0.028	0.0092	0.0021

Total cholesterol	rs9536314	G	319428	0.0042	0.0034	0.21	0.0038	0.0038	0.32	0.013	0.011	0.23
Total cholesterol	rs564481	T	319428	0.0028	0.0025	0.26	4.00E-04	0.0037	0.91	0.0094	0.0048	0.049
Triglycerides	rs1207568	A	319183	0.0016	0.0031	0.61	0.0012	0.0037	0.75	0.0066	0.0092	0.47
Triglycerides	rs9536314	G	319183	0.0018	0.0033	0.59	0.0019	0.0038	0.62	0.0035	0.011	0.75
Triglycerides	rs564481	T	319183	-0.0013	0.0025	0.59	-0.0024	0.0036	0.51	-7.00E-04	0.0047	0.88
Waist circumference	rs1207568	A	334475	0.002	0.0028	0.48	0.002	0.0033	0.53	0.0043	0.0081	0.59
Waist circumference	rs9536314	G	334475	-0.0058	0.0029	0.05	-0.0054	0.0033	0.11	-0.017	0.0097	0.073
Waist circumference	rs564481	T	334475	0.0034	0.0022	0.13	0.0025	0.0032	0.44	0.0078	0.0042	0.062

4.5.4 Effect of sex and age on *KL*-VS, G395A and C1818T associations

We sought to test whether sex or age moderate the effect of Klotho variants on cardiometabolic phenotypes by stratifying our sample by sex and by age (≥60 vs. <60), repeating the analyses separately in each sample and testing whether there is heterogeneity between the effect sizes. We found no evidence that the effect sizes between those aged ≥60 years and those aged <60 years are different (Table 4.6 & Table 4.7). We also found no evidence that the effect sizes between males and females are different (Table 4.8 & Table 4.9), with the possible exception of rs1207568 and type 2 diabetes: in men, there was a small but not statistically significant effect (beta = -0.00094 [SE = 0.019, *p* = 0.96]); in women, the effect was larger and in the opposite direction (beta = 0.062 [SE = 0.025, *p* = 0.011]); but the heterogeneity p-value of 0.045 did not pass correction for multiple testing.

Table 4.6 Age-stratified regression of cardiometabolic outcomes on *KL* variants

Log odds ratios (beta) and their respective errors (SE) when regressing the six cardiometabolic outcomes on *KL*-VS (rs9536314), G395A (rs1207568) and C1818T (rs564481) using logistic regression in participants aged ≥60 years and in participants aged <60 years *p* heterogeneity = *p*-value from chi-squared test of heterogeneity. N = sample size. A1 = effect size.

Phenotype	Variant	A1	Over 60				Under 60				*p* heterogeneity
			N	Beta	SE	*p*	N	beta	SE	*p*	
BMI	rs1207568	A	150456	0.00041	0.0047	0.93	183504	0.0034	0.0042	0.42	0.64
BMI	rs9536314	G	150456	1.80E-05	0.005	1	183504	-0.011	0.0045	0.01	0.1
BMI	rs564481	T	150456	0.00043	0.0037	0.91	183504	0.0079	0.0034	0.019	0.14
Diastolic BP	rs1207568	A	138210	0.0084	0.0048	0.079	167427	-0.00036	0.0043	0.93	0.17
Diastolic BP	rs9536314	G	138210	-0.00067	0.0052	0.9	167427	-0.00068	0.0046	0.88	1
Diastolic BP	rs564481	T	138210	-0.0036	0.0039	0.35	167427	-0.0011	0.0034	0.75	0.63
Fasting glucose	rs1207568	A	3293	-0.012	0.031	0.7	7326	0.016	0.021	0.46	0.45
Fasting glucose	rs9536314	G	3293	-0.044	0.033	0.19	7326	0.016	0.022	0.49	0.13
Fasting glucose	rs564481	T	3293	0.022	0.026	0.4	7326	-0.0015	0.017	0.93	0.45
HbA1c	rs1207568	A	133816	-0.0096	0.0049	0.052	169187	-0.011	0.0043	0.0089	0.83
HbA1c	rs9536314	G	133816	0.015	0.0053	0.006	169187	0.01	0.0046	0.027	0.48
HbA1c	rs564481	T	133816	-0.0066	0.004	0.093	169187	-0.0035	0.0034	0.3	0.55
HDL	rs1207568	A	132098	0.0038	0.0045	0.41	160399	0.0056	0.0041	0.17	0.77
HDL	rs9536314	G	132098	-0.0022	0.0049	0.65	160399	-0.0016	0.0044	0.71	0.93
HDL	rs564481	T	132098	0.00033	0.0036	0.93	160399	0.0092	0.0033	0.0049	0.069
LDL	rs1207568	A	143571	0.0029	0.0046	0.54	175266	0.0064	0.0043	0.14	0.58
LDL	rs9536314	G	143571	0.0085	0.005	0.086	175266	0.0043	0.0046	0.34	0.54
LDL	rs564481	T	143571	0.0011	0.0037	0.78	175266	-0.00058	0.0034	0.86	0.74
Lipoprotein A	rs1207568	A	114051	-8.80E-05	0.0054	0.99	139937	-0.00015	0.0048	0.98	0.99
Lipoprotein A	rs9536314	G	114051	-0.002	0.0057	0.72	139937	0.0039	0.0052	0.45	0.44
Lipoprotein A	rs564481	T	114051	-0.0057	0.0043	0.19	139937	-0.0029	0.0039	0.45	0.63
Systolic BP	rs1207568	A	138206	0.0052	0.0048	0.27	167424	-0.00014	0.0042	0.97	0.4

Systolic BP	rs9536314	G	138206	-0.0082	0.0051	0.11	167424	-0.00095	0.0044	0.83	0.28
Systolic BP	rs564481	T	138206	0.0026	0.0038	0.5	167424	0.0011	0.0033	0.74	0.77
Total cholesterol	rs1207568	A	143829	0.0053	0.0045	0.25	175599	0.0074	0.0043	0.084	0.74
Total cholesterol	rs9536314	G	143829	0.0068	0.0049	0.16	175599	0.0029	0.0045	0.52	0.56
Total cholesterol	rs564481	T	143829	0.0018	0.0036	0.62	175599	0.0029	0.0034	0.39	0.82
Triglycerides	rs1207568	A	143748	0.0051	0.0047	0.28	175435	-0.0012	0.0041	0.76	0.31
Triglycerides	rs9536314	G	143748	-0.00011	0.0051	0.98	175435	0.0038	0.0044	0.39	0.56
Triglycerides	rs564481	T	143748	-1.10E-05	0.0038	1	175435	-0.0027	0.0033	0.41	0.59
Waist circumference	rs1207568	A	150742	9.70E-05	0.0041	0.98	183733	0.0034	0.0037	0.36	0.55
Waist circumference	rs9536314	G	150742	-0.001	0.0044	0.81	183733	-0.0093	0.004	0.019	0.16
Waist circumference	rs564481	T	150742	0.0017	0.0033	0.61	183733	0.0046	0.003	0.12	0.52

Table 4.7 Age-stratified regression of cardiometabolic traits on *KL* variants

Regression coefficients (beta) and their respective errors (SE) when regressing the 11 cardiometabolic traits on *KL*-VS (rs9536314), G395A (rs1207568) and C1818T (rs564481) using linear regression in participants aged ≥60 years and in participants aged <60 years *p* heterogeneity = *p*-value from chi-squared test of heterogeneity. N = sample size.

Phenotype	Variant	A1	**Over 60**				**Under 60**				*p* heterogeneity
			N	beta	SE	*p*	N	beta	SE	*p*	
CAD	rs1207568	A	128383	-0.017	0.031	0.58	169559	0.026	0.042	0.54	0.41
CAD	rs9536314	G	128383	0.0022	0.033	0.95	169559	-0.046	0.045	0.31	0.39
CAD	rs564481	T	128383	-0.007	0.024	0.78	169559	-0.038	0.034	0.26	0.46
Cerebrovascular disease	rs1207568	A	130610	0.0086	0.035	0.8	171038	0.026	0.05	0.61	0.78
Cerebrovascular disease	rs9536314	G	130610	-0.05	0.038	0.19	171038	0.0048	0.054	0.93	0.41
Cerebrovascular disease	rs564481	T	130610	-0.0027	0.028	0.92	171038	0.016	0.04	0.7	0.7
Myocardial infarction	rs1207568	A	126541	-0.022	0.043	0.61	168872	-0.0086	0.052	0.87	0.84
Myocardial infarction	rs9536314	G	126541	0.04	0.045	0.37	168872	0.0037	0.055	0.95	0.61
Myocardial infarction	rs564481	T	126541	-0.019	0.034	0.57	168872	-0.071	0.042	0.09	0.34
Stroke	rs1207568	A	129764	0.028	0.042	0.49	170664	0.03	0.059	0.61	0.98
Stroke	rs9536314	G	129764	-0.021	0.045	0.64	170664	0.028	0.063	0.66	0.53
Stroke	rs564481	T	129764	-0.03	0.034	0.37	170664	-0.0041	0.048	0.93	0.66
Type 2 diabetes	rs1207568	A	150282	-0.02	0.019	0.3	183190	-0.029	0.025	0.24	0.77
Type 2 diabetes	rs9536314	G	150282	-0.0095	0.02	0.64	183190	-0.0041	0.026	0.87	0.87
Type 2 diabetes	rs564481	T	150282	-0.032	0.015	0.036	183190	0.0095	0.019	0.63	0.086

Table 4.8 Sex-stratified regression of cardiometabolic outcomes on *KL* variants

Log odds ratios (beta) and their respective errors (SE) when regressing the six cardiometabolic outcomes on *KL*-VS (rs9536314), G395A (rs1207568) and C1818T (rs564481) using logistic regression in female participants and in male participants *p* heterogeneity = *p*-value from chi-squared test of heterogeneity. N = sample size.

Trait	Variant	A1	Female				Male				*p* heterogeneity
			N	beta	SE	*p*	N	beta	SE	*p*	
CAD	rs1207568	A	163819	0.014	0.047	0.77	134123	-0.0086	0.029	0.77	0.68
CAD	rs9536314	G	163819	-0.016	0.05	0.76	134123	-0.014	0.031	0.66	0.97
CAD	rs564481	T	163819	0.015	0.038	0.7	134123	-0.03	0.023	0.19	0.31
Cerebrovascular disease	rs1207568	A	165269	0.014	0.043	0.74	136379	0.013	0.038	0.73	0.99
Cerebrovascular disease	rs9536314	G	165269	-0.07	0.047	0.14	136379	-0.0024	0.041	0.95	0.28
Cerebrovascular disease	rs564481	T	165269	0.021	0.035	0.55	136379	-0.01	0.031	0.74	0.51
Myocardial infarction	rs1207568	A	163145	-0.03	0.064	0.64	132268	-0.012	0.039	0.75	0.81
Myocardial infarction	rs9536314	G	163145	0.12	0.065	0.07	132268	-0.01	0.041	0.8	0.091
Myocardial infarction	rs564481	T	163145	-0.035	0.051	0.5	132268	-0.042	0.031	0.17	0.91
Premature CAD	rs1207568	A	163819	0.21	0.11	0.05	134123	-0.0033	0.065	0.96	0.095
Premature CAD	rs9536314	G	163819	-0.12	0.13	0.34	134123	-0.032	0.07	0.65	0.55
Premature CAD	rs564481	T	163819	0.037	0.091	0.68	134123	-0.072	0.052	0.17	0.3
Stroke	rs1207568	A	164698	0.019	0.053	0.72	135730	0.036	0.045	0.42	0.81
Stroke	rs9536314	G	164698	-0.034	0.057	0.55	135730	0.017	0.048	0.72	0.49
Stroke	rs564481	T	164698	-0.011	0.042	0.79	135730	-0.028	0.036	0.43	0.76
Type 2 diabetes	rs1207568	A	178906	-0.062	0.025	0.011	154566	0.00094	0.019	0.96	0.045
Type 2 diabetes	rs9536314	G	178906	-0.03	0.026	0.25	154566	0.0061	0.02	0.76	0.27
Type 2 diabetes	rs564481	T	178906	-0.027	0.019	0.17	154566	-0.0098	0.015	0.52	0.48

Table 4.9 Sex-stratified regression of cardiometabolic traits on *KL* variants

Regression coefficients (beta) and their respective errors (SE) when regressing the 11 cardiometabolic traits on *KL*-VS (rs9536314), G395A (rs1207568) and C1818T (rs564481) using linear regression in female participants and in male participants *p* heterogeneity = *p*-value from chi-squared test of heterogeneity. N = sample size. A1 = effect allele.

Trait	Variant	A1	Female				Male				*p* heterogeneity
			N	Beta	SE	*p*	N	beta	SE	*p*	
BMI	rs1207568	A	179054	-7.80E-05	0.0043	0.99	154906	0.0051	0.0046	0.27	0.41
BMI	rs9536314	G	179054	-0.007	0.0045	0.12	154906	-0.0062	0.0049	0.2	0.9
BMI	rs564481	T	179054	0.0027	0.0034	0.43	154906	0.0078	0.0037	0.034	0.31
Diastolic BP	rs1207568	A	163418	0.0034	0.0044	0.45	142219	0.0042	0.0048	0.38	0.9
Diastolic BP	rs9536314	G	163418	-0.0036	0.0047	0.45	142219	0.0026	0.0051	0.62	0.37
Diastolic BP	rs564481	T	163418	-0.00024	0.0035	0.95	142219	-0.0043	0.0038	0.26	0.43
Fasting glucose	rs1207568	A	4709	-0.013	0.026	0.6	5910	0.021	0.023	0.37	0.33
Fasting glucose	rs9536314	G	4709	-0.0076	0.028	0.79	5910	-0.0021	0.025	0.93	0.88
Fasting glucose	rs564481	T	4709	0.035	0.021	0.097	5910	-0.013	0.019	0.49	0.09
HbA1c	rs1207568	A	164895	-0.0079	0.0042	0.062	138108	-0.013	0.0048	0.0063	0.42
HbA1c	rs9536314	G	164895	0.0096	0.0045	0.034	138108	0.014	0.0051	0.005	0.52
HbA1c	rs564481	T	164895	-0.0053	0.0034	0.12	138108	-0.0043	0.0038	0.26	0.84
HDL	rs1207568	A	155604	0.0038	0.0046	0.41	136893	0.0073	0.0049	0.14	0.6
HDL	rs9536314	G	155604	-0.0044	0.0049	0.37	136893	0.0012	0.0052	0.82	0.43
HDL	rs564481	T	155604	0.0051	0.0037	0.16	136893	0.0066	0.0039	0.089	0.78
LDL	rs1207568	A	170863	0.0071	0.0043	0.098	147974	0.0013	0.0047	0.79	0.36
LDL	rs9536314	G	170863	0.0047	0.0046	0.31	147974	0.0083	0.005	0.095	0.6
LDL	rs564481	T	170863	0.0034	0.0034	0.32	147974	-0.0031	0.0037	0.41	0.2
Lipoprotein A	rs1207568	A	136626	-0.0014	0.0049	0.78	117362	0.0013	0.0053	0.8	0.71
Lipoprotein A	rs9536314	G	136626	0.0034	0.0052	0.51	117362	-0.0011	0.0056	0.84	0.56
Lipoprotein A	rs564481	T	136626	-0.0067	0.0039	0.09	117362	-0.0013	0.0042	0.75	0.35
Systolic BP	rs1207568	A	163414	0.0015	0.0041	0.71	142216	0.0031	0.0045	0.5	0.79
Systolic BP	rs9536314	G	163414	-0.0076	0.0044	0.081	142216	0.00049	0.0049	0.92	0.22

Systolic BP	rs564481	T	163414	0.0024	0.0033	0.47	142216	9.00E-04	0.0036	0.8	0.76
Total cholesterol	rs1207568	A	171141	0.0086	0.0043	0.043	148287	0.0032	0.0047	0.49	0.4
Total cholesterol	rs9536314	G	171141	0.0023	0.0046	0.62	148287	0.0078	0.005	0.12	0.42
Total cholesterol	rs564481	T	171141	0.0053	0.0034	0.12	148287	-0.00046	0.0037	0.9	0.25
Triglycerides	rs1207568	A	171059	0.0011	0.0043	0.81	148124	0.0015	0.0047	0.75	0.95
Triglycerides	rs9536314	G	171059	-0.0013	0.0046	0.78	148124	0.0052	0.005	0.3	0.34
Triglycerides	rs564481	T	171059	-0.00045	0.0034	0.89	148124	-0.0024	0.0038	0.52	0.7
Waist circumference	rs1207568	A	179276	0.00034	0.0042	0.94	155199	0.0044	0.0046	0.33	0.51
Waist circumference	rs9536314	G	179276	-0.0075	0.0045	0.099	155199	-0.0051	0.0049	0.29	0.72
Waist circumference	rs564481	T	179276	0.0036	0.0034	0.29	155199	0.0041	0.0037	0.27	0.92

4.5.5 Association between other Klotho variants and cardiometabolic traits

We next tested whether any of the remaining 211 Klotho variants are associated with any of the six cardiometabolic outcomes at a *p*-value threshold of 0.05/12. We found no associations, with and without adjusting for age and sex (Supplementary Table 4.1).

We also tested whether any of the 211 Klotho variants are associated with any of the 11 quantitative cardiometabolic traits at a *p*-value threshold of 0.05/16. We found 125 variants that are associated with HbA1c. We found no other associations, with and without adjusting for age and sex (Supplementary Table 4.2).

Since rs576674, a variant upstream of *KL*, has previously been reported to be associated with type 2 diabetes [112], we repeated our analysis for these 125 variants and adjusted for rs576674 to test whether the associations we observed are due to the variants' close proximity to rs576674 (Supplementary Table 4.3). We removed any variants that were not associated with HbA1c independently of rs576674 at a *p*-value of 0.05. Out of the 93 variants that remain, we used stepwise selection to identify 4 conditionally independent associations (Table 4.10). We attempted to replicate these 4 associations in the largest and most recent GWAS of HbA1c performed in Europeans [111] and found that the association between HbA1c and the rs495392 A allele replicates if one does not correct for the fact that an attempt is being made to replicate four, not one, associations (beta = -0.0053 [SE = 0.0022, p = 0.015]). The rs563925 association does not replicate (beta = 0 [SE = 0.002, p = 1]) and there are no results available for rs546677 and rs564823. However, rs495392 is not associated with type 2 diabetes (beta = 0.0019 [SE = 0.023, p = 0.94) in the GWAS by Cai *et al.* [113] nor is it associated with fasting glucose (beta = 0.0029 [SE = 0.0031, p = 0.35) in the GWAS by Lagou *et al*. [114], but rather is a pQTL (effect allele = A, beta = 0.16 [SE = 0.046, $p = 7.2 \times 10^{-4}$]) for haemoglobin [115].

Table 4.10 Regression of HbA1c on four *KL* variants

Regression coefficients (beta) and their respective errors (SE) when regressing HbA1c on the four conditionally independent variants. A1 = effect allele.

Variant	A1	beta	SE	*p*
rs546677	T	0.016	0.0027	1.10E-09
rs564823	C	-0.012	0.0027	1.40E-05
rs563925	A	-0.012	0.0027	2.00E-05
rs495392	A	-0.013	0.0028	5.30E-06

4.6 Discussion

The aim of the work described in this chapter was to verify the previously reported associations between *KL* variants, including the *KL*-VS haplotype, and cardiometabolic phenotypes in the UK Biobank, and to search for novel associations between *KL* genetic variants and cardiometabolic phenotypes. We found no reliable associations between *KL*-VS, G395A and C1818T and any of the phenotypes that we tested nor did we find evidence that age or sex affects these associations. We did find an association between rs495392 and HbA1c and, assuming that this is not a false positive, this may be driven by a currently unspecified non-glycaemic pathway (discussed later in this section).

The lack of reliable associations between *KL*-VS, G395A and C1818T and any of the phenotypes that we tested may not be surprising considering the fact that previously published associations between these variants and cardiometabolic phenotypes are not always consistent (likely due to the fact much of them are based on small sample studies that have not been replicated) and much of them are from individuals who do not have a European ancestry. The only study carried out using UK participants was by Freathy *et al*. [94], who did not find any associations between *KL*-VS and the cardiometabolic phenotypes that they tested, which supports the findings we report here.

One possible explanation for the apparent inconsistency between our findings and those reported previously by some other authors may be that these authors use participants that are older or younger than the samples we used. However, we did not find evidence of heterogeneity in effect sizes between those aged ≥60 years and those aged <60 years, which suggests that the effect of *KL*-VS, G395A and C1818T on cardiometabolic phenotypes is unlikely to vary with age and therefore may negate the possibility that the effect of Klotho variants on cardiometabolic phenotypes varies with age. However, it is also important to note that some authors who have reported associations between Klotho variants and cardiometabolic phenotypes used participants that were much older than the participants we used for our study [56,69,97], and the lack of evidence in our data for heterogeneity between the effect sizes of Klotho variants and cardiometabolic phenotypes may be due to the fact that almost all of the participants that we used were aged ≤69 years, so differences in effect

at older ages would not have been detected by the analyses we performed. Whilst it may be worth repeating the analyses described in this chapter when UK Biobank participants are older, it is currently not possible to conclude that age effects the relationship between Klotho variants and cardiometabolic phenotypes.

We found the rs495392 A allele decreases HbA1c concentrations but suggest that this is due to a non-glycaemic pathway. This is for two reasons: firstly, rs495392 is not associated with type 2 diabetes nor with fasting glucose; and secondly, the rs495392 A allele is reported to be associated with increased haemoglobin in the blood [115], which would have the effect of reducing the concentration of glycated haemoglobin. A possible avenue for future work could be to elucidate what pathway might explain the rs495392 association. However, it should be noted that whilst there is suggestive evidence that this association is true ($p < 0.05$), strictly speaking, the replication p-value does not pass multiple testing correction and this association may also be another false positive.

There are three main limitations to our study. Firstly, since the UKB consists of participants who are of European descent, we cannot comment on the reliability of associations between Klotho variants and cardiometabolic phenotypes reported by other authors in non-European populations. Secondly, the participants in the UKB are healthier than average, which may explain the differences between our findings and previously published reports: other studies may have less of a healthy volunteer bias. Finally, the UKB dataset may be biased for the purposes of detecting an association between Klotho variants and diseases that affect younger people, e.g. premature CAD, because these people may have either died or are otherwise unable to volunteer to participate in studies like the UKB.

In conclusion, there is insufficient evidence in the UK Biobank to support the concept that *KL* variants affect cardiometabolic outcomes or traits in British Caucasian individuals. Further follow-up testing would be required to verify the reported effects of *KL* variants on cardiometabolic outcomes and traits that are reported in very elderly individuals.

5 Klotho and Cognition

5.1 Introduction

As discussed in Section 1.1, older people have the capacity to continue to contribute socially and economically. However, this is only possible if they are able to maintain their cognitive function as they age. This chapter begins with a review of the previously-published evidence pertaining to the association between Klotho variants and cognitive function and its decline before describing the results of analyses that were carried out to verify the reported associations between *KL* variants between the *KL* genetic variants and cognitive function, including the *KL*-VS haplotype, and to search for any novel associations.

Cognition can be defined as any process that is required for an individual to be aware of their situation and to use that information to respond to it [116]. As individuals get older, memory, learning and processing speed decline; [117] often leading to reduced independence and increased reliance on families and social care.

Multiple studies have been carried out exploring the relationship between *KL* variants, cognitive function and its decline, mostly focusing on the *KL*-VS haplotype. However, previous evidence has been varied: some authors have suggested that among adults aged 70 years or more, people homozygous for V (valine) at position 352 have poorer cognitive function [118,119], but also suggests that V352 heterozygotes have better cognitive function than those who are homozygous for F (phenylalanine) at position 352 [53,118]. On the other hand, Mengel-From *et al.* [120] reported that, in Danish populations aged between 92-100 years, V352 heterozygotes had poorer cognition and Almeida *et al.* [121] reported that, among men aged 71-87 years, V352 carriers were more likely to get dementia. De Vries *et al.* [122] reported that V352 heterozygotes have a slower rate of cognitive decline, but Porter *et al.* [123] did not find any such relationship in their data.

In addition to the *KL*-VS haplotype, there are reports of associations between variants in the *KL* promoter region and cognition. Mengel-From *et al.* [120] reported that carriers of

the rs398655 C allele had better cognitive function than non-carriers and Hao *et al.* [124] reported that those who are homozygous for the G (guanine) allele at G-395A (rs1207568) have an increased risk of cognitive impairment.

At present, these conflicting reports indicate that, at a population level, the relationship between the *KL*-VS haplotype and cognitive function or cognitive decline is not particularly clear: there is, therefore, a need to explore this area further using significantly bigger sample sizes. In the work described in this chapter, we aimed to verify the reported associations between *KL* variants, including the *KL*-VS haplotype, and cognitive function in up to 335,074 UK Biobank (UKB) participants aged between 40 and 81 years, by carrying out a phenome scan of cognitive measures, including reaction time and various memory tests. We also aimed to search for novel associations between the *KL* genetic variants and cognitive function using the same approach.

5.2 Phenotypes

One of the problems with evaluating cognitive function in the UK Biobank is that dementia is usually diagnosed in people who are much older than the UKB participants [125] so there are very few cases. Additionally, individuals with dementia are: less likely to be recruited; less likely to be able to give informed consent; and less likely to be able to participate in longitudinal studies. Therefore, instead of attempting to find associations between Klotho variants and dementia directly the analyses described in this chapter used data from cognitive function tests.

Table 5.1 summarises the phenotypes relating to cognitive function (referred to as cognitive measures) that were used for our analyses. A measure of general cognitive ability, *g*, was calculated by carrying out a principal components (PC) analysis and taking the first PC (representing 40% of the variation in the four cognitive measures) as *g*. For some cognitive measures, a baseline measurement was carried out (referred to as 'Baseline', which occurred between 2006-2010) at one of 22 assessment centres as well as up to 2 follow-up measurements (referred to as 'Repeat' and 'Imaging' – the former occurred between 2012-2013, and the latter began in 2014 and is still ongoing) for a subset of participants. For Pairs Matching, there were 3 rounds; the first round had 3 pairs that the participants needed to match and the second and third rounds had 6; we

only used data from the second round because there was insufficient variation in the data from the first round, and the third round had a high level of missingness. We did not include participants in the analysis for a given cognitive function test if they abandoned the test and/or if they completed the test with a pause.

Table 5.1 Description of the cognitive measures from the UK Biobank

A description of the cognitive measures from the UK Biobank used in this study.

Category	Field ID	Description
Fluid Intelligence	20016	Number of correct answers given to 13 fluid intelligence questions within 2 minutes
Reaction Time	20023	Mean time to correctly identify matching pairs of cards (based on 12 rounds)
Numeric Memory	4282	Maximum number of digits remembered correctly
Pairs Matching	399	Number of incorrect matches in round when recalling the position of matching pairs of cards
Prospective Memory	4291	Instruction correctly recalled first time (yes/no)

Each follow-up of each measure was treated as a separate phenotype unless otherwise stated (e.g. if a measure was available for Baseline, Repeat and Imaging, then these three results are treated as three separate phenotypes). The rate of change of a cognitive measure is the change in the cognitive measure, M2-M1, divided by the age difference between Baseline (T1) and Imaging (T2), T2-T1, in years: on average, the difference between two measurements is 8.4 years. Educational attainment was based on UKB field 6138 and was coded as a binary variable (those with and without a College or University degree at baseline).

5.3 Statistical Analyses

PLINK 2.0 [66] was used to fit additive and dominant linear models between the cognitive measures and the genotypes in all individuals. This was then repeated for the following subsets of individuals: those who were aged 69 years or more at the time of performing the cognitive test (note: age^2 is not included as a covariate in analyses for this subset due to multicollinearity); and those who were Apo-ε4/ε4 (i.e. those who have a CC genotype at both rs429358 and rs7412) and those who were not.

Unless otherwise specified, all association analyses (i.e. additive and dominant linear models) were adjusted for the first 4 genetic principal components (PCs) (UKB field 22009) and the genotyping array on which the participant was genotyped. The cognitive measures and any quantitative covariates were standardised to a mean of 0 and a variance of 1 before any linear modelling was performed.

Since multiple tests were undertaken, we applied statistical correction for this. A principal component (PC) analysis showed that all 5 PCs are needed to represent >90% of the variation in the 5 cognitive measures. There were 15 independent variants based on the method described in Section 2.8. A p-value threshold of 0.05 is used and the Bonferroni-correction is applied when necessary for the appropriate number of independent tests in each case (up to 75 independent tests: 15 independent variants and 5 PCs).

5.4 Results

After QC, there were 335,028 individuals available for analysis. A summary of the sample by phenotype is provided in Table 5.2.

Table 5.2 Population characteristics by cognitive measure

Characteristics of the population by phenotype n = number of samples. SD = 1 standard deviation. All phenotypes are in their original units, except for Prospective Memory, for which the percentage of individuals who recalled the instruction correctly on the first attempt is shown.

Phenotype	n	units, mean (SD)	age, years, mean (SD)	% female	% degree
Fluid Intelligence (Baseline)	108433	6 (2.1)	57 (8)	53.3	51.7
Fluid Intelligence (Repeat)	14654	7 (2)	62 (7.3)	50.7	46.8
Fluid Intelligence (Imaging)	22981	7 (2)	63 (7.5)	50.6	47.6
g	34489	0 (1.2)	57 (8.2)	53.6	53.2
Numeric Memory (Baseline)	35484	7 (1.7)	57 (8.2)	53.6	52.7
Numeric Memory (Imaging)	15245	7 (1.4)	64 (7.4)	50.6	46.6
Pairs Matching (Baseline)	327485	4 (3.2)	57 (8)	53.7	50.8
Pairs Matching (Repeat)	14651	4 (3)	62 (7.3)	50.8	46.8
Pairs Matching (Imaging)	23040	4 (2.8)	63 (7.5)	50.6	47.6
Prospective Memory (Baseline)	110550	80.2	57 (8)	53.3	51.4
Prospective Memory (Repeat)	14766	86.2	62 (7.3)	50.7	46.7
Prospective Memory (Imaging)	23256	85.7	63 (7.5)	50.6	47.6
Reaction Time (Baseline)	333034	555 (113.2)	57 (8)	53.6	50.5
Reaction Time (Repeat)	14735	556 (109.6)	62 (7.3)	50.7	46.7
Reaction Time (Imaging)	23320	591 (108.6)	63 (7.5)	50.6	47.6

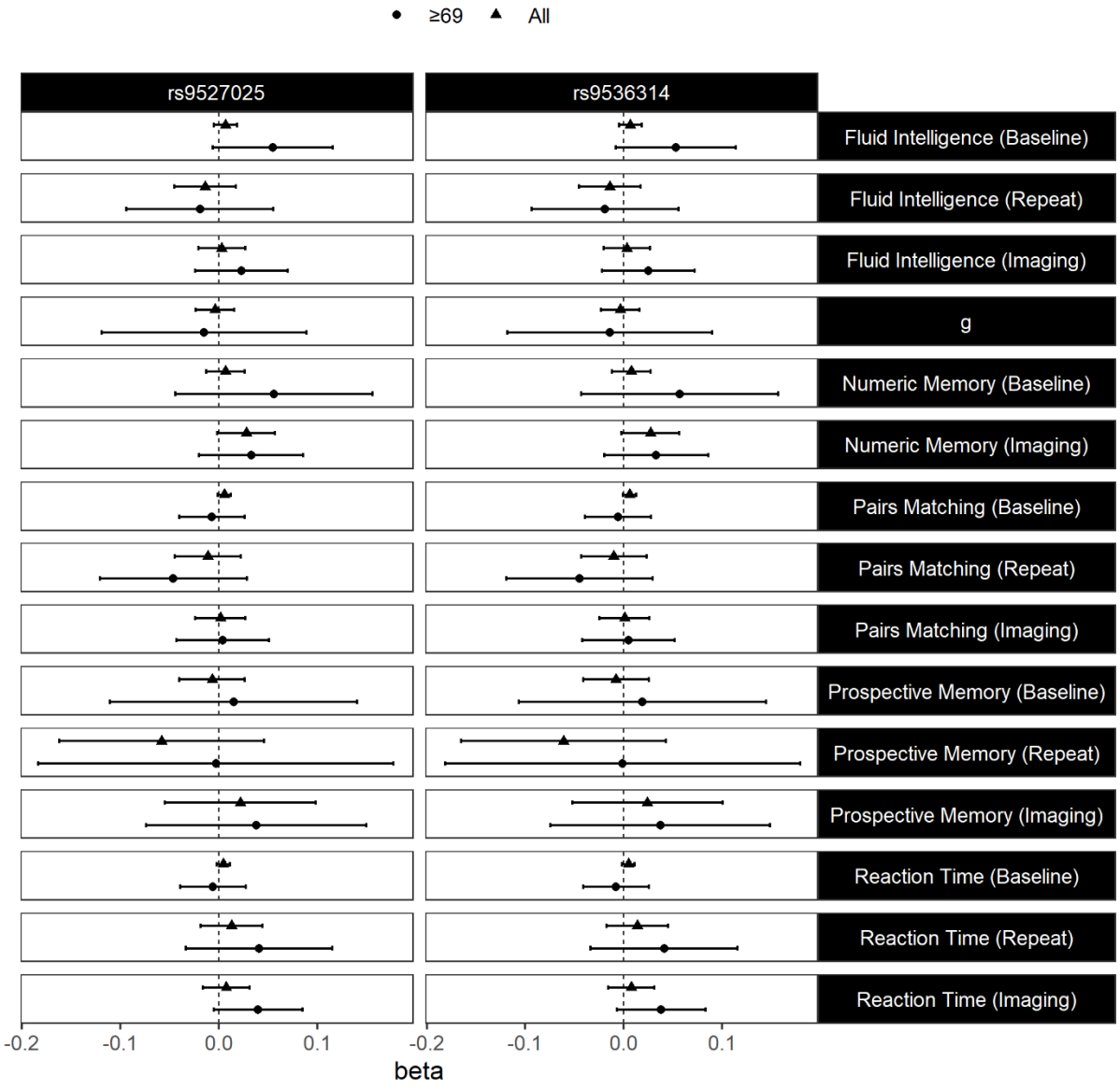
Since the 2 variants making up the *KL*-VS haplotype are well-characterised functional *KL* variants in humans, we investigated whether either of them were associated with the cognitive function measures. Neither rs9536314 nor rs9527025 were significantly associated at a p-value threshold of 0.05/5 with any of the cognitive measures (assuming an additive model) when unadjusted (Supplementary Table 5.1) and when adjusted for age, age^2, sex, and educational attainment (Supplementary Table 5.1 & Figure 5.1).

A dominant model was also tested since some previous studies compared carriers to non-carriers. This also yielded no associations (Supplementary Table 5.1).

Since our sample participants are younger than those in which associations between *KL*-VS and cognitive function have been found, we repeated our analyses in participants aged 69 years or more (n ≥ 1224). We did not find any associations (Supplementary Table 5.1 & Figure 5.1).

Figure 5.1 Regression of cognitive measures on *KL*-VS

Standardised beta coefficients with 95% Confidence Intervals when regressing cognitive measures on rs9536314 and on rs9527025 in the UK Biobank with (All) and without (≥69) including participants less than 69 years old, using an additive model, and adjusted for age, age^2, sex, and educational attainment.



Although the associations were not statistically significant, the effect size appeared to increase when excluding individuals under the age of 69 years. We, therefore, repeated the analyses but included a genotype*age interaction term to test whether the effect of *KL*-VS variants on the cognitive function measures changed with age. We found that age does not have a statistically significant effect on the relationship between *KL*-VS and any of the cognitive function measures available, at a p-value threshold of 0.01/5, adjusting for age, age^2, sex, and educational attainment (Supplementary Table 5.2).

There is evidence to suggest that *KL*-VS carrier status may influence the risk of Alzheimer's disease in APOE4 carriers. [21] We therefore divided our sample into APOE4 carriers and non-carriers and tested whether rs9536314 or rs9527025 were associated with any of the cognitive function measures, with and without adjusting for age, age^2, sex, and educational attainment. We did not find any associations at a p-value threshold of 0.05/5 in APOE4 carriers nor in non-carriers (Supplementary Table 5.3). We also included an APOE4*genotype interaction term in our model to test whether the associations between the *KL*-VS variants and the cognitive measures were affected by APOE4 carrier status, but the interaction term was not statistically significant for any of the cognitive measures tested (Supplementary Table 5.4).

We next sought to test whether rs9536314 or rs9527025 were associated with a change in any of the cognitive measures over age. For all measures except Prospective Memory, a rate of change was calculated for each participant (see Section 5.3). We found that neither rs9536314 nor rs9527025 were significantly associated at a p-value threshold of 0.05 with a change in any of the available cognitive measures over age, adjusted for the measure at baseline (M1), age$_{T1}$, age$_{T1}^2$, sex, and educational attainment (Supplementary Table 5.5).

We also tested to see if any other *KL* variants were associated at a p-value threshold of 0.05/75 with any of the available cognitive measures. The rs141113969 T allele is associated with participants being more likely to not recall the instruction on their first attempt in the Prospective Memory task during the imaging visit (log odds = 0.37 [SE = 0.1, p = 3.9×10^{-4}]), adjusting for age, age^2, sex, and educational attainment. In participants aged 69 years or older, we found that the rs2227122 T allele and the rs676046 A allele are associated with increased errors in the Pairs Matching task

during the repeat assessment centre visit (beta = 0.33 [SE = 0.094, p = 4.1×10^{-4}], and beta = 0.32 [SE = 0.092, p = 6.1×10^{-4}], respectively), adjusting for age, age^2, sex, and educational attainment. There were no further significant associations, with or without adjustment for age, age^2, sex, and educational attainment (Supplementary Table 5.1); there were also no other significant associations when excluding individuals under the age of 69 years (Supplementary Table 5.1).

We attempted to replicate the association between rs141113969 and Prospective Memory. We first removed participants who performed the Prospective Memory task during the imaging visit from the baseline sample to create an independent sample. The previously identified association did not reproduce in this sample at a p-value threshold of 0.05 (beta = 0.032 [SE = 0.053, p = 0.55]). We also found no associations at a p-value of 0.05/4 between rs141113969 and the other four cognitive measures nor between rs141113969 and *g* (Table 5.3). The age of the participants who completed the Prospective Memory task at baseline is, on average, lower in comparison to the age of the participants at the imaging visit (Table 5.2). However, there is no evidence that the relationship between rs141113969 and Prospective Memory is affected by age: firstly, the genotype*age interaction term is not significant (Supplementary Table 5.6); and secondly, a sliding window plot suggests that the association in the imaging sample is driven by younger individuals (Supplementary Figure 5.1). Therefore, the absence of an association between rs141113969 and Prospective Memory at baseline is not because the baseline sample is younger than the imaging sample. Note: the U-shape seen in Supplementary Figure 5.1 is not present in an independent baseline sample (Supplementary Figure 5.2).

Table 5.3 Regression of cognitive measures on rs141113969

Standardised coefficients (beta), standard errors (se), sample sizes (n) and p-values (p) when regressing cognitive measures on rs141113969 in the UK Biobank in participants who performed the Prospective Memory task during the imaging visit (Discovery) and those who did not (Baseline), with and without adjusting for age, age^2, sex, and educational attainment.

phenotype	n	beta	se	p
Additive model adjusted for age, age^2, sex, and educational attainment				
Fluid Intelligence (Baseline)	84433	-0.00075	0.018	0.97
g	26999	0.039	0.031	0.2
Numeric Memory (Baseline)	27474	0.065	0.031	0.036
Pairs Matching (Baseline)	249802	-0.00065	0.011	0.95
Prospective Memory (Baseline)	85332	0.032	0.053	0.55
Prospective Memory (Discovery)	21636	0.37	0.1	0.00039
Reaction Time (Baseline)	252703	0.019	0.01	0.06
Additive model adjusted for PCs and chip only				
Fluid Intelligence (Baseline)	100559	0.0024	0.018	0.9
g	32185	0.02	0.032	0.54
Numeric Memory (Baseline)	33163	0.028	0.031	0.37
Pairs Matching (Baseline)	304524	0.0077	0.01	0.45
Prospective Memory (Baseline)	102642	0.052	0.044	0.23
Prospective Memory (Discovery)	23256	0.37	0.097	0.00013
Reaction Time (Baseline)	309815	0.011	0.01	0.28
Dominant model adjusted for age, age^2, sex, and educational attainment				
Fluid Intelligence (Baseline)	84433	-0.0036	0.018	0.84
g	26999	0.034	0.032	0.28
Numeric Memory (Baseline)	27474	0.062	0.031	0.048
Pairs Matching (Baseline)	249802	-0.0016	0.011	0.89
Prospective Memory (Baseline)	85332	0.034	0.053	0.52
Prospective Memory (Discovery)	21636	0.35	0.11	0.001
Reaction Time (Baseline)	252703	0.019	0.01	0.065
Dominant model adjusted for PCs and chip only				
Fluid Intelligence (Baseline)	100559	4.00E-04	0.018	0.98
g	32185	0.014	0.032	0.66

phenotype	n	beta	se	p
Numeric Memory (Baseline)	33163	0.028	0.032	0.38
Pairs Matching (Baseline)	304524	0.0072	0.01	0.49
Prospective Memory (Baseline)	102642	0.052	0.044	0.24
Prospective Memory (Discovery)	23256	0.36	0.1	0.00029
Reaction Time (Baseline)	309815	0.011	0.01	0.3

To replicate the associations between rs2227122 and rs676046 and Pairs Matching, we removed participants who performed the Pairs Matching task during the repeat assessment visit from the imaging visit sample (and not the baseline sample because the baseline sample has no individuals aged over 70 years), tested the associations in those aged 69 years or more (adjusted for age, sex, and educational attainment), and found that neither of the previously identified associations replicate (beta = 0.052 [SE = 0.07, p = 0.46] and beta = 0.054 [SE = 0.068, p = 0.43] for rs2227122 and rs676046, respectively) (Table 5.4). We also found no associations between these two variants and the other four cognitive measures in the same sample (Table 5.4). There is no evidence of a genotype*age interaction (Supplementary Table 5.6).

Table 5.4 Regression of cognitive measures on rs2227122 and rs676046

Standardised coefficients (beta), standard errors (se), sample sizes (n) and p-values (p) when regressing cognitive measures on rs2227122 and rs676046 in the UK Biobank in participants aged ≥69 years who performed the Pairs Matching task during the repeat assessment visit (Discovery) and those who did not (Imaging), with and without adjusting for age, sex, and educational attainment.

phenotype	variant	A1	n	beta	se	p
Additive model adjusted for age, age^2, sex, and educational attainment						
Fluid Intelligence (Imaging)	rs2227122	T	4887	0.089	0.07	0.2
Numeric Memory (Imaging)	rs2227122	T	3880	0.067	0.08	0.41
Pairs Matching (Discovery)	rs2227122	T	2601	0.33	0.094	0.00041
Pairs Matching (Imaging)	rs2227122	T	4916	-0.052	0.07	0.46
Prospective Memory (Imaging)	rs2227122	T	4972	-0.26	0.17	0.13
Reaction Time (Imaging)	rs2227122	T	5019	0.053	0.069	0.45
Additive model adjusted for PCs and chip only						
Fluid Intelligence (Imaging)	rs2227122	T	4887	0.09	0.07	0.2
Numeric Memory (Imaging)	rs2227122	T	3880	0.064	0.081	0.43
Pairs Matching (Discovery)	rs2227122	T	2601	0.35	0.094	0.00023
Pairs Matching (Imaging)	rs2227122	T	4916	-0.058	0.07	0.41
Prospective Memory (Imaging)	rs2227122	T	4972	-0.28	0.17	0.1
Reaction Time (Imaging)	rs2227122	T	5019	0.055	0.07	0.43
Dominant model adjusted for age, age^2, sex, and educational attainment						
Fluid Intelligence (Imaging)	rs2227122	T	4887	0.095	0.071	0.18
Numeric Memory (Imaging)	rs2227122	T	3880	0.059	0.082	0.47
Pairs Matching (Discovery)	rs2227122	T	2601	0.33	0.094	0.00041
Pairs Matching (Imaging)	rs2227122	T	4916	-0.052	0.071	0.46
Prospective Memory (Imaging)	rs2227122	T	4972	-0.25	0.18	0.15
Reaction Time (Imaging)	rs2227122	T	5019	0.058	0.07	0.41
Dominant model adjusted for PCs and chip only						
Fluid Intelligence (Imaging)	rs2227122	T	4887	0.096	0.071	0.17
Numeric Memory (Imaging)	rs2227122	T	3880	0.058	0.082	0.48
Pairs Matching (Discovery)	rs2227122	T	2601	0.35	0.094	0.00023
Pairs Matching (Imaging)	rs2227122	T	4916	-0.058	0.071	0.41
Prospective Memory (Imaging)	rs2227122	T	4972	-0.27	0.17	0.11
Reaction Time (Imaging)	rs2227122	T	5019	0.059	0.07	0.4
Additive model adjusted for age, age^2, sex, and educational attainment						
Fluid Intelligence (Imaging)	rs676046	A	4887	0.091	0.068	0.18
Numeric Memory (Imaging)	rs676046	A	3880	0.086	0.079	0.27
Pairs Matching (Discovery)	rs676046	A	2601	0.32	0.092	0.00061
Pairs Matching (Imaging)	rs676046	A	4916	-0.054	0.068	0.43
Prospective Memory (Imaging)	rs676046	A	4972	-0.28	0.17	0.096
Reaction Time (Imaging)	rs676046	A	5019	0.053	0.067	0.43

phenotype	variant	A1	n	beta	se	p
Additive model adjusted for PCs and chip only						
Fluid Intelligence (Imaging)	rs676046	A	4887	0.09	0.068	0.19
Numeric Memory (Imaging)	rs676046	A	3880	0.081	0.079	0.31
Pairs Matching (Discovery)	rs676046	A	2601	0.33	0.093	0.00035
Pairs Matching (Imaging)	rs676046	A	4916	-0.06	0.068	0.38
Prospective Memory (Imaging)	rs676046	A	4972	-0.3	0.17	0.076
Reaction Time (Imaging)	rs676046	A	5019	0.058	0.068	0.39
Dominant model adjusted for age, age^2, sex, and educational attainment						
Fluid Intelligence (Imaging)	rs676046	A	4887	0.097	0.069	0.16
Numeric Memory (Imaging)	rs676046	A	3880	0.079	0.08	0.32
Pairs Matching (Discovery)	rs676046	A	2601	0.32	0.092	0.00061
Pairs Matching (Imaging)	rs676046	A	4916	-0.055	0.069	0.43
Prospective Memory (Imaging)	rs676046	A	4972	-0.28	0.17	0.11
Reaction Time (Imaging)	rs676046	A	5019	0.059	0.068	0.39
Dominant model adjusted for PCs and chip only						
Fluid Intelligence (Imaging)	rs676046	A	4887	0.096	0.069	0.16
Numeric Memory (Imaging)	rs676046	A	3880	0.075	0.081	0.36
Pairs Matching (Discovery)	rs676046	A	2601	0.33	0.093	0.00035
Pairs Matching (Imaging)	rs676046	A	4916	-0.061	0.069	0.38
Prospective Memory (Imaging)	rs676046	A	4972	-0.29	0.17	0.086
Reaction Time (Imaging)	rs676046	A	5019	0.062	0.069	0.37

We next sought to test whether any *KL* variants were significantly associated at a p-value threshold of 0.05/15 with a change in any of the cognitive measures over age, in the same way that *KL*-VS was tested. We did not find any statistically significant associations (Supplementary Table 5.5).

5.5 Discussion

Previous evidence suggested that *KL*-VS and other *KL* variants are associated with cognitive function during the later stages of life. Our aim was to explore these findings in a younger and much larger cohort, namely the UK Biobank, using intermediate phenotypes that can track early stages of cognitive decline [126]. We did not find evidence of a relationship between *KL*-VS and cognitive function, nor did we find any evidence that the age of an individual had a significant effect on this relationship. The very small number of associations that we found did not replicate in independent samples, nor is there any evidence of them in previously published studies, so they seem likely to be false positives. We also did not find evidence of any other *KL* variants being associated with cognitive function, or with cognitive decline.

An important point is that previous studies that have identified relationships between *KL* variants and cognition use populations that are much older (usually aged 70 years or more), whereas the population we examined is relatively young (the larger Baseline samples had a mean age of 57 years). We attempted to address this limitation by repeating our analyses, but only including individuals aged 69 years or more; we still did not find associations - perhaps because only about 4% of this subset in the imaging cohort are over 75 years of age and less than 1% are over 79 years of age. Indeed, whenever authors report an absence of statistically significant associations between *KL* variants and cognition, the mean age of the cohorts that they analyse are closer to the one we analysed. For example, Deary *et al.* [119] examined 2 cohorts and the cohort who undertook cognitive testing at age of 64 years did not show statistically significant associations between *KL*-VS and cognition. Dubal *et al.* [53] also did not find an association in one of the 3 cohorts that they analysed, and the mean age of this cohort was 63 years.

Deary *et al.* [119] provided evidence suggesting that *KL*-VS may influence cognitive decline. We did not find any evidence to support this. This may be because the difference between the repeated measurements available to us was about 8 years whereas Deary *et al.* compared the cognitive abilities of individuals first tested when aged 11 years and then at the age of 79 years. It is also important to note that whilst some authors do report relationships between *KL*-VS and cognitive decline [119,122], other authors do not find any such relationship [120,123].

The UKB dataset, despite the advantage of its size, does have biases. In particular, the participants are generally healthier than average [127]. There is evidence to suggest that the effect of *KL* variants on cognitive function/decline may be as a result of affecting the severity of a pre-existing psychopathology [128,129] and individuals suffering from early dementia, etc would be either unlikely or even unable to volunteer to participate. Furthermore, it is important to note that, in comparison to what is considered "gold standard" practice in cognitive testing, the tests undertaken by UKB participants were shorter, carried out in an uncontrolled and unsupervised environment, and did not include multiple tests to evaluate a particular aspect of cognition. This raises the question of whether or not the UKB tests are a reasonable measure of cognitive function. A recent attempt to compare UKB cognitive tests with more standard reference tests for cognitive function found that the former correlates reasonably well with the latter, suggesting that the UKB cognitive tests are a reliable, but not ideal, measure of cognitive function. The test re-test reliability (i.e. the correlation between performance on a cognitive test administered 2-4 weeks apart) of UKB tests were also moderate to high, suggesting that the UKB cognitive tests can be used to assess factors affecting cognitive decline [130]. It is also important to point out that the utility of the *g* factor is contested [131].

In conclusion, there is insufficient evidence in the UK Biobank to support the concept that *KL* variants affect cognitive function or its rate of decline in British Caucasian individuals aged between 40 and 81 years. Analysis of follow-up testing, as more participants move into old age, would be required to verify the reported effects of *KL* variants on cognitive function and decline that have been reported in very elderly individuals.

6 Klotho and Cancer

6.1 Introduction

Cancer is expected to become the leading cause of death and is claimed by some commentators to be the single most important barrier to increasing life expectancy across the world in the 21st century [132]. Thus, it is possible that Klotho might influence mortality by either affecting risk of cancer initiation or survival rate. As with other phenotypes examined in the course of this work, there are conflicting reports about the potential role of Klotho in cancer (discussed below).

This chapter begins with a brief definition of cancer, then continues with a review of the available evidence for an association between Klotho variants and cancer and, finally, describes work that uses genetic data and linked hospital records available from the UK Biobank to test the association between *KL* variants and cancer risk and survival in up to 335,024 participants.

Cancer can be defined in general terms as an abnormal and uncontrolled growth of cells. Despite the large number of differences between different types of cancer, they all share a few key characteristics [133] that are typically not exhibited by non-cancerous cells:

- The rate at which cells proliferate is partly determined by a combination of pro-growth and anti-growth signals, but cancer cells are able to both become self-sufficient in pro-growth signals (e.g. they produce their own signalling molecules and respond to these to create a positive feedback loop) [134] and are also able to adapt to avoid responding to anti-growth stimuli [135].
- In addition to being able to proliferate, cancer cells must also avoid both apoptosis (programmed cell death) [136] and avoid becoming senescent (normal cells cannot double more than approximately 70 times, which is referred to as their Hayflick limit) [137].

- Finally, as tumours become more advanced, they are able to generate their own network of blood vessels (angiogenesis) [138] and are able to migrate and invade other tissues (metastasis) [139].

Carcinogenesis is a multistep process, beginning with a genetic aberration that may, amongst other things, result in a tumour suppressor gene becoming inactivated. Evidence from several authors implies that the Klotho (*KL*) gene may be a tumour suppressor and it has also been shown that decreased *KL* expression is associated with poorer prognosis in a variety of different cancers [140]. For example, increased Klotho expression inhibited colony formation in MCF-7 and MDA-MB-231 cells, which are both breast cancer cell lines [141].

Furthermore, some authors have reported associations between *KL* variants and cancer in non-European populations. Kim *et al.* [142] found that rs3752472 is associated with prostate cancer risk in a sample of Korean men. Liu *et al.* [143] and Kamal *et al.* [144] found that carriers of the rs1207568 A allele were more likely to get colorectal cancer in a Chinese and in an Egyptian population, respectively. However, neither of these reports presented any evidence for an association between colorectal cancer risk and rs564481 [143,144], though it could be argued that rs1207568 is a promoter variant and may affect Klotho expression, whereas there is no conclusive evidence that rs564481 is a functional variant.

To date, the relationship between genetic variants at the Klotho locus and cancer risk and survival has not been studied in a sufficiently large European sample. Here we use genetic data and linked hospital records available from the UK Biobank to test the association between *KL* variants and the risk of different types of cancer as well as the effect of *KL* variants on the survival chances of participants diagnosed with these different types of cancer.

6.2 Data sources

The UK Biobank was used as the discovery sample (see Section 2.2). The replication dataset for melanoma survival consisted of 858 non-Hispanic white patients with cutaneous melanoma recruited between March 1998 and August 2008 [145].

6.3 Phenotypes

The UKB provides separate data fields related to participants' cancer diagnoses from national cancer registries. UKB field 40006 lists all of the cancers that a participant has been diagnosed with since records are available. UKB field 40005 lists the corresponding dates on which they received those diagnoses. No self-reported data is used for the analyses described in this chapter. For the survival analyses, data from death registries was used, namely date of death (UKB field 40000) and cause of death (UKB field 40001).

The phenotype definitions provided here are illustrated using malignant melanoma as an example. Supplementary Table 6.1 lists the ICD-10 codes used to define the grouped cancer categories (similar to the process described in Section 4.3.2). Participants with diagnoses of in-situ carcinomas or benign neoplasms (ICD-10 code beginning with D in UKB field 40006), but without the corresponding malignant neoplasm ICD-10 code were set to missing (see Supplementary Table 6.1).

Malignant melanoma: participants with one or more occurrences of ICD-10 code C43 under UKB field 40006 were considered cases; and all other participants were considered controls. The same definition was used for all other cancers, substituting the ICD-10 code for malignant melanoma with the code for each of the other cancers in turn.

To carry out survival analysis using the aforementioned malignant melanoma cases, participants with a diagnosis of malignant melanoma before baseline (diagnosis dates obtained from UKB field 40005) were excluded. In addition, participants with a diagnosis of any other cancer before their malignant melanoma diagnosis were also excluded. The survival time in days for each participant was calculated as the number of days between the date on which they were first diagnosed with malignant melanoma and the censor date, where the censor date was the date of death for participants who had died (UKB field 40000) or 26 April 2020 for participants who had survived until this date. A participant whose primary cause of death (UKB field 40001) was malignant melanoma (ICD-10 code C43) was considered an event. The age of first diagnosis for a given individual is the age at which that individual was first diagnosed with malignant

melanoma. The same procedure was used for all other cancers, substituting the ICD-10 code for malignant melanoma with the code for each of the other cancers in turn.

6.4 Statistical analyses

We used R 4.0.2 [75] for all analyses, unless stated otherwise. We used PLINK 2.0 [66] to regress each of the cancer outcomes on each of the *KL* variants using logistic regression, assuming an additive genetic model, and adjusting for age at baseline, sex, the array on which the participant was genotyped and the first four principal components for the genetic variability of the genome (UKB field 22009).

We omitted all cancers with <661 cases based on a power calculation [146], which showed that 661 cases are required to detect a relative risk of 1.2 at a p-value of 0.05 with 80% power assuming a MAF of 0.16 (because rs9536314 has a MAF of 0.16) and a control:case ratio of 5.62:1 (because this is the overall cancer_free:cancer ratio).

We used the coxph() function from the survival 3.2.3 package in R [147] to fit Cox proportional hazards models to assess the association between *KL* variants and cancer survival. We used sex, the age at which the participant was first diagnosed with the cancer, the first four genetic principal components and the genotyping array as covariates. Only cancers included in the aforementioned logistic regression analyses were included.

As we carried out multiple tests, we applied a statistical correction. There are two independent genetic variants (see Section 2.8). We tested both individual cancers and cancer groups. We assume that each cancer group and any cancer that does not fit within a group are independent, so p-values generated from the analyses described in this chapter are corrected for up to 20 independent tests (threshold = 2.5×10^{-3}).

6.5 Results

After the QC described in Chapter 2, there were 335,024 participants remaining. Their characteristics are summarised in Table 6.1 and the number of cases for the different types of cancer analysed are presented in Table 6.2.

Table 6.1 Age and sex distribution by cancer status in the UK Biobank

Number of participants with (cases) and without (controls) one or more cancer diagnoses and the average age and percentage female of these groups.

	All	Cancer	Controls
N	335,024	50,597	284,427
age in years (SD)	57 (8)	60 (6.7)	56 (8)
% female	53.6	51.7	53.9

Table 6.2 Prevalence of selected cancers in the UK Biobank

Number of participants diagnosed at least once with specific cancer, excluding those who did not pass QC. Note that the All category includes cancer with too few individual cases as per the power calculation.

Cancer	cases	controls
Head and Neck	1026	332434
Digestive	6208	327146
Upper GI Tract	1074	332424
Lower GI Tract	4227	329124
Lower Respiratory	1903	331627
Female Reproductive	2611	173602
Male Reproductive	7276	147162
Breast	9614	167871
Melanoma	2681	329938
Skin (not melanoma)	17227	315733
Bladder	919	332182
Prostate	6818	147647
C18 Colon	2605	330835
C20 Rectum	1236	332199
C34 Lung	1900	331630
C54 Uterus	1253	177658
C56 Ovary	842	178076
C64 Kidney	964	332570
C83 Non-Hodgkin's lymphoma	767	332754
All	50580	277692

6.5.1 *KL* variants and cancer risk

Since the two variants making up the *KL*-VS haplotype are well-characterised functional *KL* variants in humans, we investigated whether or not they were associated with cancer risk. We did not find evidence of an association between rs9536314 and the available cancers in our sample ($p > 0.05$), with or without adjusting for age and sex (Supplementary Table 6.2). We repeated our analyses using dominant, recessive, and heterozygous models, but still did not find any significant associations (Supplementary Table 6.2).

We next tested the association between the remaining 213 *KL* variants and cancer risk (Supplementary Table 6.2). However, we did not find sufficient evidence of an association between any of these variants and the available cancers in our sample ($p > 2.5 \times 10^{-3}$).

6.5.2 *KL* variants and cancer survival

We tested the association between the 214 *KL* variants and cancer survival (Supplementary Table 6.3). We found that, in our sample, the rs71436501 G allele and the rs78425544 G allele are independently associated with reduced melanoma survival (HR = 2.54 [95% CIs 1.65 to 3.92, $p = 2.3 \times 10^{-5}$] and HR = 2.32 [95% CIs 1.49 to 3.62, $p = 2.1 \times 10^{-4}$], respectively) and the rs1207570 C allele and the rs438793 G allele were associated with reduced head and neck cancer (HNC) survival (HR = 1.83 [95% CIs 1.28 to 2.61, $p = 8.7 \times 10^{-4}$] and HR = 1.74 [95% CIs 1.22 to 2.49, $p = 2.4 \times 10^{-3}$], respectively).

However, the association between the rs71436501 G and the rs78425544 G alleles and melanoma survival was not present in the replication sample (HR = 0.44 [95% CIs 0.14 to 1.42, p = 0.17] and HR = 0.44 [95% CIs 0.14 to 1.44, p = 0.17], respectively). The associations between rs1207570 and rs438793 and HNC survival are yet to be replicated.

6.6 Discussion

Previous evidence suggested that *KL* variants were associated with colorectal cancer risk in non-European populations, but similar work had not been carried out in a large European sample. We did not find sufficient evidence to suggest that Klotho variants are associated with the types of cancer that we tested.

Despite the association between *KL*-VS (rs9536314) and numerous other traits (see previous chapters), we did not find evidence of an association between rs9536314 and cancer risk or survival, albeit only in the types of cancer that we tested. Furthermore, we did not find an association between rs1207568 and colorectal cancer, although the previous groups who reported this association found it in Chinese [143] and Egyptian populations [144]. However, Liu *et al.* [143] and Kamal *et al.* [144] were not able to find evidence of an association between rs564481 and colorectal cancer and neither did we. The rs3752472 variant that is reportedly associated with prostate cancer [142] is not available in the UKB (because it is a rare variant in individuals of European ancestry).

We tested whether Klotho variants were associated with surviving the types of cancer that we tested. We found associations between rs71436501 and rs78425544 and melanoma survival. However, we could not replicate these associations, so they are likely to be false positives. The replication sample we used was of European ancestry and, although it was smaller than the discovery sample, the direction of effect was opposite to the one observed in the discovery sample, so statistical power is a secondary issue in this case. It should be noted that replication performed in the replication sample was adjusted for age and sex and included only individuals who had early stage melanoma, which is similar to what was done for the discovery analyses.

The associations between rs1207570 and rs438793 and HNC survival are yet to be replicated and should be treated with caution for now. As part of preparing the work presented in this chapter for publication, corresponding authors on papers that have GWAS data for HNC patients have been contacted and it is hoped that this will allow for an attempt to be made to replicate these associations.

Although the UKB cohort is healthier than average [127], the number of cancer cases we use is similar to and sometimes even larger than those used by other groups to provide evidence of associations between *KL* variants and cancer risk, albeit in non-Caucasian populations. Therefore, the number of cases is unlikely to be the reason why we did not find any associations with cancer risk. However, as this sample has a White British ancestry, our work should not be taken as a comment on the reliability of the associations between *KL* variants and colorectal risk reported in Chinese [143] and Egyptian populations [144].

In conclusion, we do not find evidence at a population genetics level to suggest that Klotho is associated with cancer risk in individuals with a European ancestry. Despite evidence suggesting that Klotho expression is associated with tumour progression, the associations we found between *KL* variants and melanoma survival and HNC survival fail to replicate and are yet to be replicated, respectively, and should be considered as potential false positives for now.

It is important to note that the genetics of a cancerous cell are very different to that of a normal cell and common germline variants obtained from genotyping non-cancerous cells are by no means sufficient to capture this. Therefore, given the considerable evidence of Klotho expression affecting cancer prognosis, Klotho should not necessarily be dismissed as a target for anti-cancer research.

7 *KL* gene-based and *KL-FGF23-FGFR1* gene set analysis

7.1 Introduction

So far in this thesis, all analyses have focused on the relationship between individual variants in and around the *KL* gene locus and age-related phenotypes. In this chapter, using summary statistics from previously published genome-wide association studies of parental lifespan, myocardial infarction, cancer, COVID-19 risk and Alzheimer's disease, the association between variants located around genes that encode the proteins that make up the Klotho signalling complex (namely, *KL*, *FGF23* and *FGFR1*) are analysed as a whole as opposed to separately.

7.2 Methods

7.2.1 GWAS summary statistics

As discussed in Chapter 1, a genome-wide association study (GWAS) is a hypothesis-free approach that typically examines millions of common variants across the genome for an association with a phenotype of interest. One advantage of more recent GWASs is their large sample sizes, which is particularly beneficial for case-control phenotypes because the UK Biobank cohort is healthier than average which results in fewer cases and therefore less statistical power. Table 7.1 summarises the studies from which summary statistics were used to carry out the analyses in this chapter. It should be noted here that all summary statistics that were used were generated from participants with a European ancestry.

Table 7.1 GWASs used for gene-based and gene set analyses

Authors, phenotypes and sample sizes for GWASs used for gene-based and gene set analyses

Author	Phenotype	Sample size
Timmers *et al.* [29].	Parental lifespan	500,193 European ancestry offspring
Hartiala *et al.* [148].	Myocardial infarction	14,825 European ancestry cases and 380,970 European ancestry controls
Michailidou *et al.* [149].	Breast cancer	122,977 cases and 105,974 controls
Schumacher *et al.* [150].	Prostate cancer	79,148 cases and 61,106 controls
Wang *et al.* [151].	Lung cancer	11,348 cases and 15,861 controls
Kunkle *et al.* [152].	Alzheimer's disease	21,982 cases and 41,944 controls
COVID-19 Host Genetics Initiative [153]	COVID-19 risk (C2)	122,616 cases and 2,475,240 controls[a]

[a]obtained from summary statistics file because information not available for release 7

7.2.2 GWAS of colorectal cancer in the UK Biobank

A GWAS of colorectal cancer was performed in the UK Biobank data. In addition to the QC described in Chapter 2, all variants with a MAF < 0.05 were excluded. Individuals who had ICD-10 codes C18 (Malignant neoplasm of colon) and/or C19 (Malignant neoplasm of rectosigmoid junction) and/or C20 (Malignant neoplasm of rectum) in field 40006 (Type of cancer: ICD10) were coded as cases. If an individual was not a case but had ICD-10 codes D010 (Carcinoma in situ, Colon) and/or D011 (Carcinoma in situ, Rectosigmoid junction) and/or D012 (Carcinoma in situ, Rectum) in field 40006 (Type of cancer: ICD10), they were set to missing. After QC, there were 4,107 cases and 329,266 controls remaining. All analyses were logistic regressions performed using PLINK 2.0 [66] and adjusted for age, sex, the first four genetic principal components and the genotyping chip.

7.2.3 Selection of genes for pathway analysis

As discussed in Chapter 1, FGF23–FGFR1c–Klotho complexes form dimers that allow FGF23 signalling to take place. However, the proteins involved in the downstream signalling pathways are by no means exclusive to Klotho-dependent FGF23 signalling. Therefore, in an attempt to avoid possible false positives that might arise due to genes being involved in large numbers of signalling pathways, the set of genes used for the gene set analyses described in this chapter are restricted to *KL*, *FGF23* and *FGFR1*.

7.2.4 Gene-based and gene set analysis using MAGMA

Models run using MAGMA

MAGMA v1.10 [154] was used to perform gene-set analyses on the GWAS summary statistics referred to in sections 7.2.1 and 7.2.2. In the first step, a gene-based analysis is performed which combines the p-values for associations at a gene locus into a mean test statistic, which results in an overall p-value for that gene. In the second step, a model is run that tests whether a gene being in a gene set is associated with its Z-score (obtained from the p-values generated in the first step); a statistically significant p-value

at this second step would suggest that the genes in the gene set of interest are more likely to be associated with the phenotype in comparison to genes not the gene set.

1000 Genomes

At this point, it is important to note that LD between the variants is also taken into account in the first step of the analysis, which requires a reference data set. The analyses described in this chapter use 1000 Genomes data [155,156] downloaded from: https://www.cog-genomics.org/plink/2.0/resources. In addition to removing related and non-European samples (this information is provided as part of the download), the following QC was performed: samples with a missingness of >97% and/or extremes of heterozygosity (±3 standard deviations) were removed; and variants with a missingness of >97% and/or a MAF of <5% and/or not in HWE ($p < 10^{-6}$) were removed. The remaining samples and variants were used as the reference data set which MAGMA uses to estimate LD between variants.

Assigning variants to genes

In order to perform a gene-based analysis, MAGMA requires information to link together variants with genes. A variant was assigned to a gene if its position was within a gene's transcription start and end sites. In addition, variants that were found to kidney eQTLs (expression quantitative traits loci, which are variants that are associated with the expression of a gene) for *KL* [156] were also considered to be associated with *KL* even if they did not fall within the transcription start and end sites. Supplementary File 7.1 lists which variants were assigned to which genes.

Table 7.2 Kidney eQTLs for *KL*

Variants associated with *KL* expression in kidney tubules [156]

variant	other allele	effect allele	beta	SE	p-value
rs7324259	G	T	-0.173	0.027	3.09E-10
rs6561643	A	T	-0.139	0.022	1.15E-09
rs9535729	T	G	-0.117	0.023	8.36E-07
rs9596553	G	A	-0.116	0.023	9.04E-07
rs731089	T	C	-0.113	0.023	1.60E-06
rs2348264	T	A	-0.113	0.023	1.63E-06

MULTIPLE TESTING

Since the gene-based and gene set analyses were performed on eight different GWASs, an association is only considered statistically significant if the p-value is less than 0.05/8 = 0.00625. Please see Section 2.8.

7.3 Results

7.3.1 Gene-based results for *KL*

A gene-based analysis was performed for all genes for each of the eight GWASs. Table 7.3 shows the results for *KL*. There are no statistically significant associations ($p > 0.00625$) between *KL* and the eight traits that were tested when using a gene-based approach. The analysis was repeated using a publicly available GWAS of phosphate (https://gwas.mrcieu.ac.uk/datasets/ukb-d-30810_raw/) as a positive control and a significant association for *KL* was found ($p = 5.09 \times 10^{-8}$).

Table 7.3 Gene-based results for *KL*

Z-scores and p-values from gene-based analyses for *KL* nVariants = number of variants.

trait	nVariants	z-statistic	p-value
Breast cancer	178	0.779	0.218
Alzheimer's disease	163	-0.087	0.535
Colorectal cancer	188	-0.595	0.724
COVID-19 risk (C2)	186	1.386	0.083
Myocardial infarction	188	1.764	0.039
Lung cancer	172	-0.408	0.658
Prostate cancer	188	0.617	0.269
Parental lifespan	175	0.656	0.256

7.3.2 Gene set results for *KL-FGF23-FGFR1*

A gene set analysis was performed for a Klotho-specific gene set (containing *KL*, *FGF23* and *FGFR1*) for each of the eight GWASs. There are no statistically significant associations between the *KL-FGF23-FGFR1* gene set and the eight traits that were tested. Please see Table 7.4 for the results. The analysis was repeated using a publicly available GWAS of phosphate (https://gwas.mrcieu.ac.uk/datasets/ukb-d-30810_raw/) as a positive control and a significant association for *KL-FGF23-FGFR1* was found ($p = 3.36 \times 10^{-6}$).

Table 7.4 Gene set results for *KL*-FGF23-FGFR1

Regression coefficients and their standard errors from a gene set analysis performed using MAGMA. Analyses are adjusted for gene size, gene density, inverse mac, log(gene size), log(gene density) and log(inverse mac) mac = minor allele count.

trait	beta	SE	p-value
Breast cancer	0.811	0.505	0.054
Alzheimer's disease	-1.074	0.472	0.989
Colorectal cancer	0.445	0.472	0.172
COVID-19 risk (C2)	0.596	0.448	0.092
Myocardial infarction	0.232	0.507	0.323
Lung cancer	0.270	0.467	0.281
Prostate cancer[a]	-0.243	0.705	0.635
Parental lifespan	-0.031	0.502	0.525

[a]Note that gene-based results for *FGF23* not available for prostate cancer, so gene set result is based on *KL* and *FGFR1* only.

7.4 Discussion

In this chapter, gene-based and gene set approaches were used to evaluate the association between eight ageing-related traits and *KL* and *KL-FGF23-FGFR1*, respectively. Neither of these approaches yielded statistically significant results.

To date, the closest analysis that has been performed with which the analyses described in this chapter can be compared is a gene-based PheWAS on exome sequencing data from the UK Biobank [157]. The results from this PheWAS analysis also show a nominally significant association between *KL* and myocardial infarction, but this association is no longer significant when corrected for multiple testing, which is consistent with the findings presented in this chapter.

To the best of one's knowledge, there are no other studies that perform gene set analyses using a *KL-FGF23-FGFR1* gene set. However, since an association was found between the *KL-FGF23-FGFR1* gene set and phosphate, the approach and the selection of genes is likely to be valid. Therefore, the lack of association between the *KL-FGF23-FGFR1* gene set and the age-related traits tested in this chapter is likely to be a robust finding, although independent replication is always ideal.

A limitation of both the gene-based analysis and, by extension, the gene set analysis (since it relies on the results of the gene-based analysis) is determining which variants correspond to which gene. The analyses in this chapter assume that variants located between the transcription start and end sites for a gene are associated with that gene. However, this may not always be the case: one issue with this assumption is that there may be variants that are located upstream of a gene in its promoter region that might affect its expression (and adding a fixed window may not always solve this issue because the distance between the transcription start site and these promoter variants is likely to vary); and another issue is that variants might even be found in the transcribed areas of two different genes in cases where gene loci overlap (i.e. one gene on the positive strand and the other on the negative strand). Another approach could be to select variants that are either exons of a gene or are known eQTLs for that same and group these together, but it is important to note that there are variants that affect the expression of more than one gene.

In conclusion, combining together variants in a gene-based analysis or combining together genes in a gene set analysis does not provide evidence for *KL* being associated with age-related phenotypes.

8 Discussion

The aim of this thesis was to evaluate the Klotho gene as a target for anti-ageing research. In this chapter, the overall findings from this thesis are summarised in light of this aim as well as the strengths and limitations of the approaches and data before suggestions for possible avenues for future research are presented.

8.1 Summary of results

Throughout this thesis, the genetic variants that make up the *KL*-VS haplotype are tested for associations with age-related phenotypes. This is because reports of associations between *KL* variants and age-related phenotypes (reviewed in previous chapters) almost always test *KL*-VS, probably because the variants that make up the *KL*-VS haplotype may well be functional variants [52-55]. However, despite testing *KL*-VS separately from other variants to avoid having to adjust the p-values for multiple testing, no reliable associations were found between *KL*-VS and longevity, cardiometabolic phenotypes, cognitive function and decline and cancer risk and survival.

We also tested other common genetic variants at the Klotho gene locus for an association with longevity, cardiometabolic phenotypes, cognitive function and decline and cancer risk and survival. There was preliminary evidence for the following associations: rs2283368 and rs9536338 with longevity; rs495392 with HbA1c; rs141113969, rs2227122 and rs676046 with memory; and rs71436501 and rs78425544 with melanoma survival. However, none of these associations could be adequately replicated.

8.2 Strengths and limitations of the UK Biobank

8.2.1 Sample size

As discussed in previous chapters, the evidence for Klotho variants being associated with age-related phenotypes is contradictory. This is likely due to the fact that these studies are based on small sample sizes, which means that the findings reported in

these studies tend to have a large margin of error and are also not replicated because there is a tendency to favour work that describes a positive result over work that describes negative or null results [28]. In order to address the problem of small sample size, data from the UK Biobank, a prospective cohort study of half a million participants, is used. A larger sample size results in improved statistical power.

However, the UK Biobank has a few disadvantages. These include healthy volunteer bias, a limited age range and participants predominantly who are of White British ancestry.

8.2.2 Age range

One of the recruitment criteria used [158] for the UK Biobank study was that participants were aged between 40 and 69 years of age. This poses two problems. Firstly, the majority of the participants are considered to be middle-aged, at least if one accepts the MeSH definition of middle-aged, i.e. an adult aged 45-64 years [82]. Given that the aim of this thesis is to evaluate Klotho as a target for anti-ageing therapy, it could be argued that using a dataset that predominantly consists of middle-aged participants is not optimal.

Secondly, if one assumes that the results of previous studies are not false positives, then it is important to consider the possibility the effect of Klotho variants on age-related phenotypes may vary with age and it is possible that the age range of the participants in the UKB may not capture the variation. There is evidence for this with respect to longevity [70], cognition [119] and coronary artery disease [90], though one cannot exclude the possibility that these pieces of evidence are false positives.

For cognition and cardiometabolic phenotypes, the UK Biobank participants were stratified in an attempt to detect any potential interaction between age and Klotho genotype. However, this approach cannot completely overcome the problem of age-specific effects. Firstly, stratification reduces sample size and this in turn reduces statistical power, though it should be noted that the sample size of the UKB is so large that even stratification does not reduce the sample size to anywhere near that of previous CGA studies on Klotho. Secondly, and more importantly, effects reported in

very old participants or relatively young participants cannot be observed in the UKB simply because there are no participants within those age ranges.

8.2.3 Healthy volunteer bias

It is well known that UK Biobank participants are socio-economically better off, have healthier lifestyles and are less likely to have a medical condition [85]. This means that any findings may not necessarily be generalisable to the UK population.

In addition, healthy volunteer bias has further implications beyond generalisability with respect to searching for associations between Klotho variants and age-related phenotypes. When performing case-control analyses, the number of cases affects the statistical power of these analyses, so a sample that is healthier than average will have fewer cases and, therefore, will have lower statistical power. However, it should be noted that the number of cases present in the data used to carry out the analyses described in previous chapters contains many more cases compared to the data used by previous authors for their CGA studies.

Beyond any potential issues of statistical power, the UKB study is unlikely to include participants who are severely unwell. An example of how this might affect the findings in this thesis is discussed in Chapter 5, where two studies [128,129] are cited that suggest that the effect of *KL* variants on cognitive function/decline may be as a result of affecting the severity of a pre-existing psychopathology and, if this is true, it could be argued that individuals suffering from early dementia, etc would be either unlikely or even unable to volunteer to participate in the UKB study and, therefore, an effect, if present, is less likely to be detected. Another example of how healthy volunteer bias might affect the findings in this thesis is discussed in Chapter 4, where previous reports of Klotho variants affecting the risk of premature CAD are cited [90,91] and, if one assumes that these reports are not false positives, it could be argued that individuals who are suffering from premature CAD would also be unlikely to volunteer for the UKB study and, therefore, an effect, if present, is again less likely to be detected.

8.2.4 Phenotype quality

The UKB is one of the first large scale studies that includes deep phenotyping data. This makes it suitable for phenome-wide association studies, which is where one attempts to find an association between a genetic variant and multiple phenotypes.

However, a large sample size can sometimes mean that one is restricted with respect to the types of data that can be collected. For example, whilst the tests used to assess the cognitive function are reportedly adequate [130], they are still not optimal in comparison to a clinical assessment [159].

In addition, several of the analyses presented throughout this thesis use hospital records. Even if one assumes that the diagnoses are accurate, it is definitely possible to argue that only those participants with particularly severe medical conditions would be diagnosed in a hospital environment. The most obvious impact of this is a fewer number of cases. However, diseases associated with ageing tend to be progressive in their nature and the initial symptoms are likely to be diagnosed in a General Practice (GP) setting. Whilst GP records of UKB participants are now becoming available, it was not possible to incorporate them into the analyses described in this thesis due to time constraints, though it may not have made a difference in any case (see below).

On the other hand, several of the analyses in this thesis use quantitative intermediate measures, such as blood pressure or reaction time. These measures are associated with diseases outcomes that tend to occur later in life and, if Klotho is indeed a suitable target for anti-ageing research, it could be argued that the analyses presented in this thesis should have shown evidence of a reliable association between Klotho variants and these quantitative intermediate phenotypes. Since there appears to be insufficient evidence of association between Klotho variants and the quantitative intermediate phenotypes analysed in this thesis, it could be argued that incorporating GP records would not have produced substantially different findings.

8.3 Overall findings and recommendations

The aim of this thesis was to evaluate Klotho as a candidate for anti-ageing research using a candidate gene association approach. The hypothesis was that if Klotho was indeed a suitable candidate for anti-ageing research, then genetic variants located in and around the Klotho gene locus should be associated with ageing as measured through longevity or be associated with diseases and phenotypes that are correlated with increased age.

Overall, the analyses presented in this thesis show that there is either no evidence of an association between Klotho variants and age-related phenotypes at all and, whenever an association was found, it did not replicate. Therefore, at present, Klotho does not appear to be a suitable target for anti-ageing research. However, in light of the limitations presented in Section 7.2 and in light of published evidence, there are some possible avenues that could be considered with respect to Klotho.

The first set of recommendations continues to focus on the candidate gene association approach. Firstly, many of the previous candidate gene association studies have been carried out using data from individuals who are not of European ancestry. Since large scale biobanks using data from individuals of non-European ancestry are likely to be created, such as the China Kadoorie Biobank [160], it may be worth performing a phenome-wide association study in these non-European biobanks. Secondly, since UK Biobank participants will continue to be followed up for many years to come, it may also be worth exploring the possibility that Klotho variants have longitudinal effects and/or that the effects of Klotho manifest in very old age.

There are numerous published reports which suggest that serum Klotho, which is produced by enzymatic cleavage of full-length Klotho, may be associated with age-related phenotypes and disease progression. These reports are based on correlations and, therefore, are susceptible to confounder bias. As proteomics continue to become more common and are likely to be applied to large scale studies [161], this opens up the possibility of both trying to replicate any associations that have been previously reported between serum Klotho and age-related phenotypes but, if one combines genetic data with proteomic data, it may be possible to perform Mendelian

randomisation studies [162] to determine whether the previously reported associations between serum Klotho and age-related phenotypes are causal.

8.4 Conclusion

In conclusion, the lack of evidence of a reliable association between Klotho variants and age-related phenotypes suggests that associations reported by previous authors who used a similar approach are likely to be false positives and, on this basis, the Klotho gene does not appear to be a suitable target for anti-ageing research. The work presented in this thesis also highlights the importance of carrying out replication studies.

References

1. [Anonymous]. Overview of the UK population - Office for National Statistics. . 2021. Available:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/january2021#the-uks-population-is-ageing.

2. Office for National Statistics. Living longer: how our population is changing and why it matters . . 2018. Available:
https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/livinglongerhowourpopulationischangingandwhyitmatters/2018-08-13.

3. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. . 2020. Available: http://ghdx.healthdata.org/gbd-results-tool.

4. Iparraguirre J. The Economic Contribution of Older People in the United Kingdom – An
Update to 2017. . 2017. Available: https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/active-communities/the_economic_contribution_of_older_-people_-update_-to_-2017.pdf.

5. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. Cell. 2013;153: 1194-1217 doi: 10.1016/j.cell.2013.05.039.

6. Kennedy BK, Monnat RJ, Kudlow BA. Werner and Hutchinson-Gilford progeria syndromes: mechanistic basis of human progeroid diseases. Nature reviews. Molecular cell biology. 2007;8: 394-404 doi: 10.1038/nrm2161.

7. Guo Y, Kim C, Ahmad S, Zhang J, Mao Y. CENP-E–dependent BubR1 autophosphorylation enhances chromosome alignment and the mitotic checkpoint. The Journal of Cell Biology. 2012;198: 205-217 doi: 10.1083/jcb.201202152.

8. Baker DJ, Dawlaty MM, Wijshake T, Jeganathan KB, Malureanu L, Ree JHv *et al*. Increased expression of BubR1 protects against aneuploidy and cancer and extends healthy lifespan. Nature cell biology. 2013;15: 96-102 doi: 10.1038/ncb2643.

9. Rossiello F, Jurk D, Passos JF, d'Adda di Fagagna F. Telomere dysfunction in ageing and age-related diseases. Nature cell biology. 2022;24: 135-147 doi: 10.1038/s41556-022-00842-x.

10. Morris BJ, Willcox BJ, Donlon TA. Genetic and epigenetic regulation of human aging and longevity. Biochimica et biophysica acta. Molecular basis of disease. 2019;1865: 1718-1744 doi: 10.1016/j.bbadis.2018.08.039.

11. Meller A, Shalgi R. The aging proteostasis decline: From nematode to human. Experimental cell research. 2021;399: 112474 doi: 10.1016/j.yexcr.2021.112474.

12. Pignatti C, D'Adamo S, Stefanelli C, Flamigni F, Cetrullo S. Nutrients and Pathways that Regulate Health Span and Life Span. Geriatrics (Basel). 2020;5: 95 doi: 10.3390/geriatrics5040095.

13. Lima T, Li TY, Mottis A, Auwerx J. Pleiotropic effects of mitochondria in aging. Nat Aging. 2022;2: 199-213 doi: 10.1038/s43587-022-00191-2.

14. Di Micco R, Krizhanovsky V, Baker D, d'Adda di Fagagna F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. Nature reviews. Molecular cell biology. 2021;22: 75-95 doi: 10.1038/s41580-020-00314-w.

15. Oh J, Lee YD, Wagers AJ. Stem cell aging: mechanisms, regulators and therapeutic opportunities. Nature medicine. 2014;20: 870-880 doi: 10.1038/nm.3651.

16. Fafián-Labora JA, O'Loghlen A. Classical and Nonclassical Intercellular Communication in Senescence and Ageing. Trends in cell biology. 2020;30: 628-639 doi: 10.1016/j.tcb.2020.05.003.

17. Qiu C, Fratiglioni L. Aging without Dementia is Achievable: Current Evidence from Epidemiological Research. Journal of Alzheimer's Disease. 2018;62: 933-942 doi: 10.3233/JAD-171037.

18. Pavlidis N, Stanta G, Audisio RA. Cancer prevalence and mortality in centenarians: A systematic review. Critical Reviews in Oncology/Hematology. 2012;83: 145-152 doi: 10.1016/j.critrevonc.2011.09.007.

19. Ismail K, Nussbaum L, Sebastiani P, Andersen S, Perls T, Barzilai N *et al*. Compression of Morbidity Is Observed Across Cohorts with Exceptional Longevity. Journal of the American Geriatrics Society. 2016;64: 1583-1591 doi: 10.1111/jgs.14222.

20. Sebastiani P, Perls TT. The Genetics of Extreme Longevity: Lessons from the New England Centenarian Study. Front Genet. 2012;3: 277 doi: 10.3389/fgene.2012.00277.

21. Herskind AM, McGue M, Holm NV, Sørensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet. 1996;97: 319-323 doi: 10.1007/BF02185763.

22. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S *et al*. Role of Genes and Environments for Explaining Alzheimer Disease. Archives of general psychiatry. 2006;63: 168-174 doi: 10.1001/archpsyc.63.2.168.

23. Willemsen G, Ward KJ, Bell CG, Christensen K, Bowden J, Dalgård C *et al*. The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. Twin Res Hum Genet. 2015;18: 762-771 doi: 10.1017/thg.2015.83.

24. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. J Intern Med. 2002;252: 247-254 doi: 10.1046/j.1365-2796.2002.01029.x.

25. Uno M, Nishida E. Lifespan-regulating genes in C elegans npj aging and mechanisms of disease. 2016;2: 16010 doi: 10.1038/npjamd.2016.10.

26. Piper MDW, Partridge L. Drosophila as a model for ageing. Biochimica et biophysica acta. Molecular basis of disease. 2018;1864: 2707-2717 doi: 10.1016/j.bbadis.2017.09.016.

27. Yuan R, Peters LL, Paigen B. Mice as a Mammalian Model for Research on the Genetics of Aging. ILAR journal. 2011;52: 4-15 doi: 10.1093/ilar.52.1.4.

28. Ioannidis JPA. This I believe in genetics: discovery can be a nuisance, replication is science, implementation matters. Front Genet. 2013;4 doi: 10.3389/fgene.2013.00033.

29. Timmers PR, Mounier N, Lall K, Fischer K, Ning Z, Feng X *et al*. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances eLife. 2019;8 doi: 10.7554/eLife.39856.

30. Wright KM, Rand KA, Kermany A, Noto K, Curtis D, Garrigan D *et al*. A Prospective Analysis of Genetic Variants Associated with Human Lifespan. G3 (Bethesda, Md.). 2019;9: 2863-2878 doi: 10.1534/g3.119.400448.

31. Melzer D, Pilling LC, Ferrucci L. The genetics of human ageing. Nature Reviews Genetics. 2020;21: 88-101 doi: 10.1038/s41576-019-0183-6.

32. Kuro-o M, Matsumura Y, Aizawa H, Kawaguchi H, Suga T, Utsugi T *et al*. Mutation of the mouse klotho gene leads to a syndrome resembling ageing. Nature. 1997;390: 45-51 doi: 10.1038/36285.

33. Urakawa I, Yamazaki Y, Shimada T, Iijima K, Hasegawa H, Okawa K *et al*. Klotho converts canonical FGF receptor into a specific receptor for FGF23. Nature. 2006;444: 770-774 doi: 10.1038/nature05315.

34. Goetz R, Nakada Y, Hu MC, Kurosu H, Wang L, Nakatani T *et al*. Isolated C-Terminal Tail of FGF23 Alleviates Hypophosphatemia by Inhibiting FGF23-FGFR-Klotho Complex Formation. Proceedings of the National Academy of Sciences of the United States of America. 2010;107: 407-412 doi: 10.1073/pnas.0902006107.

35. Ohnishi M, Razzaque MS. Dietary and genetic evidence for phosphate toxicity accelerating mammalian aging. FASEB J. 2010;24: 3562-3571 doi: 10.1096/fj.09-152488.

36. BERGWITZ C. Dietary phosphate modifies lifespan in Drosophila. Nephrology, dialysis, transplantation. 2012;27: 3399-3406 doi: 10.1093/ndt/gfs362.

37. Wang Y, Sun Z. Current Understanding of Klotho. Ageing Res Rev. 2009;8: 43-51 doi: 10.1016/j.arr.2008.10.002.

38. Riancho JA, Valero C, Hernandez JL, Ortiz F, Zarrabeitia A, Alonso MA *et al*. Association of the F352V variant of the Klotho gene with bone mineral density. Biogerontology. 2007;8: 121-127 doi: 10.1007/s10522-006-9039-5 [doi].

39. Kinoshita S, Kawai M. The FGF23/KLOTHO Regulatory Network and Its Roles in Human Disorders. Vitamins and hormones. 2016;101: 151-174 doi: 10.1016/bs.vh.2016.02.001.

40. Chen G, Liu Y, Goetz R, Fu L, Jayaraman S, Hu M *et al*. α-Klotho is a non-enzymatic molecular scaffold for FGF23 hormone signalling. Nature. 2018;553: 461-466 doi: 10.1038/nature25451.

41. Agoro R, Ni P, Noonan ML, White KE. Osteocytic FGF23 and Its Kidney Function. Frontiers in endocrinology (Lausanne). 2020;11: 592 doi: 10.3389/fendo.2020.00592.

42. Portale AA, Zhang MYH, David V, Martin A, Jiao Y, Gu W *et al*. Characterization of FGF23-Dependent Egr-1 Cistrome in the Mouse Renal Proximal Tubule. PLOS ONE. 2015;10: e0142924 doi: 10.1371/journal.pone.0142924.

43. Küng CJ, Haykir B, Schnitzbauer U, Egli-Spichtig D, Hernando N, Wagner CA. Fibroblast growth factor 23 leads to endolysosomal routing of the renal phosphate cotransporters NaPi-IIa and NaPi-IIc in vivo. American Journal of Physiology-Renal Physiology. 2021;321: F785-F798 doi: 10.1152/ajprenal.00250.2021.

44. Murer H, Hernando N, Forster I, Biber J. Proximal Tubular Phosphate Reabsorption: Molecular Mechanisms. Physiological Reviews. 2000;80: 1373-1409 doi: 10.1152/physrev.2000.80.4.1373.

45. Chanakul A, Zhang MYH, Louw A, Armbrecht HJ, Miller WL, Portale AA *et al*. FGF-23 Regulates CYP27B1 Transcription in the Kidney and in Extra-Renal Tissues. PLOS ONE. 2013;8: e72816 doi: 10.1371/journal.pone.0072816.

46. Dai B, David V, Alshayeb HM, Showkat A, Gyamlani G, Horst RL *et al*. Assessment of 24,25(OH)2D levels does not support FGF23-mediated catabolism of vitamin D metabolites. Kidney International. 2012;82: 1061-1070 doi: 10.1038/ki.2012.222.

47. Andrukhova O, Slavic S, Smorodchenko A, Zeitz U, Shalhoub V, Lanske B *et al*. FGF23 regulates renal sodium handling and blood pressure. EMBO Molecular Medicine. 2014;6: 744-759 doi: 10.1002/emmm.201303716.

48. Andrukhova O, Smorodchenko A, Egerbacher M, Streicher C, Zeitz U, Goetz R *et al*. FGF23 promotes renal calcium reabsorption through the TRPV5 channel. The EMBO Journal. 2014;33: 229-246 doi: 10.1002/embj.201284188.

49. Chen C, Podvin S, Gillespie E, Leeman SE, Abraham CR. Insulin stimulates the cleavage and release of the extracellular domain of Klotho by ADAM10 and ADAM17. Proc Natl Acad Sci U S A. 2007;104: 19796-19801 doi: 10.1073/pnas.0709805104.

50. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31: 3555-3557 doi: 10.1093/bioinformatics/btv402.

51. Arking DE, Krebsova A, Macek MS, Macek M, Arking A, Mian IS *et al*. Association of human aging with a functional variant of klotho. Proc Natl Acad Sci U S A. 2002;99: 856-861 doi: 10.1073/pnas.022484299 [doi].

52. Tucker Zhou TB, King GD, Chen C, Abraham CR. Biochemical and functional characterization of the klotho-VS polymorphism implicated in aging and disease risk. J Biol Chem. 2013;288: 36302-36311 doi: 10.1074/jbc.M113.490052.

53. Dubal DB, Yokoyama JS, Zhu L, Broestl L, Worden K, Wang D *et al*. Life extension factor klotho enhances cognition. Cell Rep. 2014;7: 1065-1076 doi: 10.1016/j.celrep.2014.03.076.

54. Yokoyama JS, Marx G, Brown JA, Bonham LW, Wang D, Coppola G *et al*. Systemic klotho is associated with KLOTHO variation and predicts intrinsic cortical connectivity in healthy human aging. Brain Imaging Behav. 2017;11: 391-400 doi: 10.1007/s11682-016-9598-2.

55. Zimmermann M, Köhler L, Kovarova M, Lerche S, Schulte C, Wurster I *et al*. The longevity gene Klotho and its cerebrospinal fluid protein profiles as a modifier for Parkinson´s disease. Eur J Neurol. 2021 doi: 10.1111/ene.14733.

56. Paula RS, Souza VC, Machado-Silva W, Almeida BRS, Daros AC, Gomes L *et al*. Serum Klotho (but not haplotypes) associate with the post-myocardial infarction status of older adults. Clinics (Sao Paulo). 2016;71: 725-732 doi: 10.6061/clinics/2016(12)09.

57. Wang H, Wang Z, Xu Q, Zhang Y, Si L, Li X *et al*. A potential regulatory single nucleotide polymorphism in the promoter of the Klotho gene may be associated with essential hypertension in the Chinese Han population. Clinica Chimica Acta. 2010;411: 386-390 doi: 10.1016/j.cca.2009.12.004.

58. Donate-Correa J, Martín-Núñez E, Martínez-Sanz R, Muros-de-Fuentes M, Mora-Fernández C, Pérez-Delgado N *et al*. Influence of Klotho gene polymorphisms on vascular gene expression and its relationship to cardiovascular disease. Journal of Cellular and Molecular Medicine. 2016;20: 128-133 doi: https://doi.org/10.1111/jcmm.12710.

59. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M *et al*. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med. 2015;3: 769-781 doi: 10.1016/S2213-2600(15)00283-0.

60. Applied Biosystems. UKB_WCSGAX: UK Biobank 500K Samples Genotyping Data Generation by the Affymetrix1 Research Services Laboratory. . 2017.

61. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K *et al*. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562: 203-3 doi: 10.1038/s41586-018-0579-z.

62. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26: 2867-2873 doi: 10.1093/bioinformatics/btq559.

63. O'Connell J, Sharp K, Shrine N, Wain L, Hall I, Tobin M *et al*. Haplotype estimation for biobank scale datasets. Nat Genet. 2016;48: 817-820 doi: 10.1038/ng.3583.

64 van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg MV, Slagboom PE, de Bakker, Paul I. W *et al*. Population-specific genotype imputations using minimac or IMPUTE2. Nat Protoc. 2015;10: 1285-1296 doi: 10.1038/nprot.2015.077.

65. Schmidt AF, Finan C. Linear regression and the normality assumption. Journal of Clinical Epidemiology. 2018;98: 146-151 doi: 10.1016/j.jclinepi.2017.12.006.

66. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4: 7 doi: 10.1186/s13742-015-0047-8.

67. Borras C, Ingles M, Mas-Bargues C, Dromant M, Sanz-Ros J, Román-Domínguez A *et al*. Centenarians: An excellent example of resilience for successful ageing. Mechanisms of ageing and development. 2020;186: 111199 doi: 10.1016/j.mad.2019.111199.

68. Arking DE, Krebsova A, Macek M, Macek M, Arking A, Mian IS *et al*. Association of human aging with a functional variant of klotho. PNAS. 2002;99: 856-861.

69. Arking DE, Atzmon G, Arking A, Barzilai N, Dietz HC. Association between a functional variant of the KLOTHO gene and high-density lipoprotein cholesterol, blood pressure, stroke, and longevity. Circ Res. 2005;96: 412-418 doi: 01.RES.0000157171.04054.30 [pii].

70. Invidia L, Salvioli S, Altilia S, Pierini M, Panourgia MP, Monti D *et al*. The frequency of Klotho *KL*-VS polymorphism in a large Italian population, from young subjects to centenarians, suggests the presence of specific time windows for its effect. Biogerontology. 2010;11: 67-73 doi: 10.1007/s10522-009-9229-z.

71. Novelli V, Viviani Anselmi C, Roncarati R, Guffanti G, Malovini A, Piluso G *et al*. Lack of replication of genetic associations with human longevity. Biogerontology. 2008;9: 85-92 doi: 10.1007/s10522-007-9116-4.

72. Flachsbart F, Franke A, Kleindorp R, Caliebe A, Blanché H, Schreiber S *et al*. Investigation of genetic susceptibility factors for human longevity - a targeted nonsynonymous SNP study. Mutat Res. 2010;694: 13-19 doi: 10.1016/j.mrfmmm.2010.08.006.

73. Collerton J, Davies K, Jagger C, Kingston A, Bond J, Eccles M *et al*. Health and disease in 85 year olds: Baseline findings from the Newcastle 85+ cohort study. British Medical Journal. 2009 doi: 10.1136/bmj.b4904.

74. Deelen J, Beekman M, Uh H, Broer L, Ayers *KL*, Tan Q *et al*. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Hum Mol Genet. 2014;23: 4420-4432 doi: 10.1093/hmg/ddu139.

75. R Core Team. R: A language and environment for statistical computing https://www.R-project.org/. 2019.

76. [Anonymous]. QCTOOL v2. . Available: https://www.well.ox.ac.uk/~gav/qctool_v2/.

77. [Anonymous]. GTOOL. . Available: https://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html.

78. Szumilas M. Explaining Odds Ratios. J Can Acad Child Adolesc Psychiatry. 2010;19: 227-229.

79. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evidence-Based Mental Health. 2019;22: 153-160 doi: 10.1136/ebmental-2019-300117.

80. Office for National Statistics. National life tables – life expectancy in the UK: 2017 to 2019. . 2020. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2017to2019#life-expectancy-at-birth-in-the-uk.

81. Suzman R, Riley MW. Introducing the "Oldest Old". Milbank Memorial Fund quarterly. Health and society. 1985;63: 175-186 doi: 10.2307/3349879.

82. MeSH. Middle Aged. . 2004. Available: https://www.ncbi.nlm.nih.gov/mesh/68008875.

83. Invidia L, Salvioli S, Altilia S, Pierini M, Panourgia M, Monti D *et al*. The frequency of Klotho *KL*-VS polymorphism in a large Italian population, from young subjects to centenarians, suggests the presence of specific time windows for its effect. Biogerontology. 2010;11: 67-73 doi: 10.1007/s10522-009-9229-z.

84. Aburto JM, Schöley J, Kashnitsky I, Zhang L, Rahal C, Missov TI *et al*. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. International journal of epidemiology. 2021 doi: 10.1093/ije/dyab207.

85. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T *et al*. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017;186: 1026-1034 doi: 10.1093/aje/kwx246.

86. Fries JF. Aging, Natural Death, and the Compression of Morbidity. New England Journal of Medicine. 1980;303: 130-135 doi: 10.1056/NEJM198007173030304.

87. Chopra AK. Metabolic Syndrome or Insulin Resistance: Evolution, Controversies and Association With Cardiovascular Disease Risk. Indian Journal of Clinical Cardiology. 2020;1: 77-85 doi: 10.1177/2632463620935030.

88. Tune JD, Goodwill AG, Sassoon DJ, Mather KJ. Cardiovascular Consequences of Metabolic Syndrome. Transl Res. 2017;183: 57-70 doi: 10.1016/j.trsl.2017.01.001.

89. Office for National Statistics. Leading causes of death, UK: 2001 to 2018

Registered leading causes of death by age, sex and country. . 2020. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/leadingcausesofdeathuk/latest.

90. Arking DE, Becker DM, Yanek LR, Fallin D, Judge DP, Moy TF *et al*. KLOTHO Allele Status and the Risk of Early-Onset Occult Coronary Artery Disease. Am J Hum Genet. 2003;72: 1154-1161.

91. Low AF, O'Donnell CJ, Kathiresan S, Everett B, Chae CU, Shaw SY *et al*. Aging syndrome genes and premature coronary artery disease. BMC Med Genet. 2005;6: 38 doi: 10.1186/1471-2350-6-38.

92. Majumdar V, Nagaraja D, Christopher R. Association of the functional *KL*-VS variant of Klotho gene with early-onset ischemic stroke. Biochemical and Biophysical Research Communications. 2010;403: 412-416 doi: 10.1016/j.bbrc.2010.11.045.

93. Pereira RMR, Freitas TQ, Franco AS, Takayama L, Caparbo VF, Domiciano DS *et al*. KLOTHO polymorphisms and age-related outcomes in community-dwelling older subjects: The São Paulo Ageing & Health (SPAH) Study. Scientific Reports. 2020;10: 1-8 doi: 10.1038/s41598-020-65441-y.

94. Freathy RM, Weedon MN, Melzer D, Shields B, Hitman GA, Walker M *et al*. The functional "*KL*-VS" variant of KLOTHO is not associated with type 2 diabetes in 5028 UK Caucasians. BMC Med Genet. 2006;7: 51 doi: 10.1186/1471-2350-7-51.

95. Nzietchueng R, El Shamieh S, Benachour H, Labat C, Herbeth B, Ndiaye NC *et al*. Klotho *KL*-VS genotype is involved in blood pressure regulation. Clinica Chimica Acta. 2011;412: 1773-1777 doi: 10.1016/j.cca.2011.05.032.

96. Majumdar V, Christopher R. Association of exonic variants of Klotho with metabolic syndrome in Asian Indians. Clinica Chimica Acta. 2011;412: 1116-1121 doi: 10.1016/j.cca.2011.02.034.

97. Paroni G, Seripa D, Panza F, Addante F, Copetti M, D'Onofrio G *et al*. Klotho locus, metabolic traits, and serum hemoglobin in hospitalized older patients: a genetic association analysis. Age (Dordr). 2012;34: 949-968 doi: 10.1007/s11357-011-9273-x.

98. Imamura A, Okumura K, Ogawa Y, Murakami R, Torigoe M, Numaguchi Y *et al*. Klotho gene polymorphism may be a genetic risk factor for atherosclerotic coronary

artery disease but not for vasospastic angina in Japanese. Clinica Chimica Acta. 2006;371: 66-70 doi: 10.1016/j.cca.2006.02.021.

99. Jo S, Kim S, Choi YJ, Joo N, Cho G, Choi S *et al*. KLOTHO gene polymorphism is associated with coronary artery stenosis but not with coronary calcification in a Korean population. Int Heart J. 2009;50: 23-32 doi: 10.1536/ihj.50.23.

100. Kim Y, Kim J, Nam YJ, Kong M, Kim YJ, Yu K *et al*. Klotho is a genetic risk factor for ischemic stroke caused by cardioembolism in Korean females. Neuroscience Letters. 2006;407: 189-194 doi: 10.1016/j.neulet.2006.08.039.

101. Rhee E, Oh K, Lee W, Kim S, Jung C, Kim B *et al*. The differential effects of age on the association of KLOTHO gene polymorphisms with coronary artery disease. Metabolism - Clinical and Experimental. 2006;55: 1344-1351 doi: 10.1016/j.metabol.2006.05.020.

102. Rhee EJ, Oh KW, Yun EJ, Jung CH, Lee WY, Kim SW *et al*. Relationship between polymorphisms G395A in promoter and C1818T in exon 4 of the KLOTHO gene with glucose metabolism and cardiovascular risk factors in Korean women. J Endocrinol Invest. 2006;29: 613-618 doi: 10.1007/BF03344160.

103. Gao LL, Ding X, Xie DM, Yang M, Dong BR. G-395A polymorphism in the promoter region of the KLOTHO gene and hypertension among elderly (90 years and older) Chinese individuals. Genet Mol Res. 2015;14: 15444-15452 doi: 10.4238/2015.November.30.22.

104. Shimoyama Y, Nishio K, Hamajima N, Niwa T. KLOTHO gene polymorphisms G-395A and C1818T are associated with lipid and glucose metabolism, bone mineral density and systolic blood pressure in Japanese healthy subjects. Clinica Chimica Acta. 2009;406: 134-138 doi: 10.1016/j.cca.2009.06.011.

105. Kuzan A. Toxicity of advanced glycation end products (Review). Biomedical Reports. 2021;14: 1-8 doi: 10.3892/br.2021.1422.

106. Eyth E, Naik R. Hemoglobin A1C. In: Anonymous StatPearls. Treasure Island (FL): StatPearls Publishing; 2021.

107. NICE. Familial hypercholesterolaemia: identification and management. . 2019. Available: https://www.nice.org.uk/guidance/cg71/chapter/recommendations#case-finding-and-diagnosis.

108. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L *et al*. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009;41: 666-676 doi: 10.1038/ng.361.

109. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Statistics in Medicine. 2005;24: 2911-2935 doi: 10.1002/sim.2165.

110. Hebbali A. Tools for Building OLS Regression Models [R package olsrr version 0.5.3]. . 2020. Available: https://CRAN.R-project.org/package=olsrr.

111. Wheeler E, Leong A, Liu C, Hivert M, Strawbridge RJ, Podmore C *et al*. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLOS Medicine. 2017;14: e1002383 doi: 10.1371/journal.pmed.1002383.

112. Zhao W, Rasheed A, Tikkanen E, Lee J, Butterworth AS, Howson JMM *et al*. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. Nature Genetics. 2017;49: 1450-1457 doi: 10.1038/ng.3943.

113. Cai L, Wheeler E, Kerrison ND, Luan J, Deloukas P, Franks PW *et al*. Genome-wide association analysis of type 2 diabetes in the EPIC-InterAct study. Scientific Data. 2020;7: 1-6 doi: 10.1038/s41597-020-00716-7.

114. Lagou V, Mägi R, Hottenga J, Grallert H, Perry JRB, Bouatia-Naji N *et al*. Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. Nat Commun. 2021;12: 24 doi: 10.1038/s41467-020-19366-9.

115. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J *et al*. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017;8 doi: 10.1038/ncomms14357.

116. Hopper L, Brosnan S. Primate Cognition. Nature Education Knowledge. 2012;5.

117. Tun PA, Lachman ME. Age differences in reaction time and attention in a national telephone sample of adults: education, sex, and task complexity matter. Dev Psychol. 2008;44: 1421-1429 doi: 10.1037/a0012845.

118. Yokoyama JS, Sturm VE, Bonham LW, Klein E, Arfanakis K, Yu L *et al*. Variation in longevity gene KLOTHO is associated with greater cortical volumes. Ann Clin Transl Neurol. 2015;2: 215-230 doi: 10.1002/acn3.161.

119. Deary IJ, Harris SE, Fox HC, Hayward C, Wright AF, Starr JM *et al*. KLOTHO genotype and cognitive ability in childhood and old age in the same individuals. Neuroscience Letters. 2005;378: 22-27 doi: 10.1016/j.neulet.2004.12.005.

120. Mengel-From J, Soerensen M, Nygaard M, McGue M, Christensen K, Christiansen L. Genetic Variants in KLOTHO Associate With Cognitive Function in the Oldest Old Group. The journals of gerontology. Series A, Biological sciences and medical sciences. 2016;71: 1151-1159 doi: 10.1093/gerona/glv163.

121. Almeida OP, Morar B, Hankey GJ, Yeap BB, Golledge J, Jablensky A *et al*. Longevity Klotho gene polymorphism and the risk of dementia in older men. Maturitas. 2017;101: 1-5 doi: 10.1016/j.maturitas.2017.04.005.

122 de Vries CF, Staff RT, Harris SE, Chapko D, Williams DS, Reichert P *et al*. Klotho, APOEε4, cognitive ability, brain size, atrophy and survival : A study in the Aberdeen Birth Cohort of 1936. . 2017. Available: http://hdl.handle.net/2164/10140.

123. Porter T, Burnham SC, Milicic L, Savage G, Maruff P, Lim YY *et al*. Klotho allele status is not associated with Aβ and APOE ε4-related cognitive decline in preclinical Alzheimer's disease. Neurobiol Aging. 2019;76: 162-165 doi: 10.1016/j.neurobiolaging.2018.12.014.

124. Hao Q, Ding X, Gao L, Yang M, Dong B. G-395A polymorphism in the promoter region of the KLOTHO gene associates with reduced cognitive impairment among the oldest old. AGE. 2016;38: 1-8 doi: 10.1007/s11357-015-9869-7.

125. Fishman E. Risk of Developing Dementia at Older Ages in the United States. Demography. 2017;54: 1897-1919 doi: 10.1007/s13524-017-0598-7.

126. Novotný JS, Gonzalez-Rivas JP, Medina-Inojosa JR, Lopez-Jimenez F, Geda YE, Stokin GB. Investigating cognition in midlife. Alzheimer's & Dementia: Translational Research & Clinical Interventions. 2021;7: e12234 doi: 10.1002/trc2.12234.

127. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Allen NE. OP41 The representativeness of the UK Biobank cohort on a range of sociodemographic, physical, lifestyle and health-related characteristics. J Epidemiol Community Health. 2016;70: A26-A26 doi: 10.1136/jech-2016-208064.41.

128. Morar B, Badcock JC, Phillips M, Almeida OP, Jablensky A. The longevity gene Klotho is differentially associated with cognition in subtypes of schizophrenia. Schizophr Res. 2018;193: 348-353 doi: 10.1016/j.schres.2017.06.054.

129. Erickson CM, Schultz SA, Oh JM, Darst BF, Ma Y, Norton D *et al*. KLOTHO heterozygosity attenuates APOE4-related amyloid burden in preclinical AD. Neurology. 2019;92: e1878-e1889 doi: 10.1212/WNL.0000000000007323.

130. Fawns-Ritchie C, Deary IJ. Reliability and validity of the UK Biobank cognitive tests. PLOS ONE. 2020;15 doi: 10.1371/journal.pone.0231627.

131. Kell HJ, Lang JWB. The Great Debate: General Ability and Specific Abilities in the Prediction of Important Outcomes. J Intell. 2018;6 doi: 10.3390/jintelligence6030039.

132. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018;68: 394-424 doi: 10.3322/caac.21492.

133. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell. 2000;100: 57-70.

134. Ungefroren H. Autocrine TGF-β in Cancer: Review of the Literature and Caveats in Experimental Analysis. International journal of molecular sciences. 2021;22: 977 doi: 10.3390/ijms22020977.

135. Amin, A. R. M. Ruhul, Karpowicz PA, Carey TE, Arbiser J, Nahta R, Chen ZG *et al*. Evasion of anti-growth signaling: A key step in tumorigenesis and potential target for treatment and prophylaxis by natural compounds. Seminars in cancer biology. 2015;35: S55-S77 doi: 10.1016/j.semcancer.2015.02.005.

136. Fernald K, Kurokawa M. Evading apoptosis in cancer. Trends in cell biology. 2013;23: 620-633 doi: 10.1016/j.tcb.2013.07.006.

137. Dimri GP. What has senescence got to do with cancer? Cancer cell. 2005;7: 505-512 doi: 10.1016/j.ccr.2005.05.025.

138. Lugano R, Ramachandran M, Dimberg A. Tumor angiogenesis: causes, consequences, challenges and opportunities. Cellular and molecular life sciences : CMLS. 2020;77: 1745-1770 doi: 10.1007/s00018-019-03351-7.

139. Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: a hallmark of cancer revisited. Signal transduction and targeted therapy. 2020;5: 28 doi: 10.1038/s41392-020-0134-x.

140. Sachdeva A, Gouge J, Kontovounisios C, Nikolaou S, Ashworth A, Lim K *et al*. Klotho and the Treatment of Human Malignancies. Cancers. 2020;12: 1665 doi: 10.3390/cancers12061665.

141. Ligumsky H, Rubinek T, Merenbakh-Lamin K, Yeheskel A, Sertchook R, Shahmoon S *et al*. Tumor suppressor activity of Klotho in breast cancer is revealed by structure-function analysis. Molecular cancer research. 2015;13: 1398-1407 doi: 10.1158/1541-7786.MCR-15-0141.

142. Kim HJ, Lee J, Lee SY, Cheong HS, Kye Y, Kim W *et al*. The association between *KL* polymorphism and prostate cancer risk in Korean patients. Mol Biol Rep. 2014;41: 7595-7606 doi: 10.1007/s11033-014-3647-y.

143. Liu C, Cui W, Wang L, Yan L, Ruan X, Liu Y *et al*. Klotho gene polymorphisms are related to colorectal cancer susceptibility. International journal of clinical and experimental pathology. 2015;8: 7446-7449.

144. Kamal A, Salama M, Kamal A, Mohsen A, Siam I. Klotho (rs1207568 and rs564481) gene variants and colorectal cancer risk. Turk J Gastroenterol. 2020;31: 497-502 doi: 10.5152/tjg.2020.19235.

145. Li H, Wang Y, Liu H, Shi Q, Xu Y, Wu W *et al*. Genetic variants in the integrin signaling pathway genes predict cutaneous melanoma survival. International journal of cancer. 2017;140: 1270-1279 doi: 10.1002/ijc.30545.

146. PURCELL S, CHERNY SS, SHAM PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics (Oxford, England). 2003;19: 149-150 doi: 10.1093/bioinformatics/19.1.149.

147. Therneau TM. A Package for Survival Analysis in R. . 2020. Available: https://CRAN.R-project.org/package=survival.

148. Hartiala JA, Han Y, Jia Q, Hilser JR, Huang P, Gukasyan J *et al*. Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. European heart journal. 2021;42: 919-933 doi: 10.1093/eurheartj/ehaa1040.

149. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S *et al*. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551: 92-94 doi: 10.1038/nature24284.

150. Schumacher FR, Olama AAA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ *et al*. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nature Genetics. 2018;50: 928-936 doi: 10.1038/s41588-018-0142-8.

151. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P *et al*. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nature Genetics. 2014;46: 736-741 doi: 10.1038/ng.3002.

152. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC *et al*. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019;51: 414-430 doi: 10.1038/s41588-019-0358-2.

153. The COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. European journal of human genetics : EJHG. 2020;28: 715-718 doi: 10.1038/s41431-020-0636-6.

154. Leeuw CAd, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLOS Computational Biology. 2015;11: e1004219 doi: 10.1371/journal.pcbi.1004219.

155. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526: 68-74 doi: 10.1038/nature15393.

156. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J *et al*. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526: 75-81 doi: 10.1038/nature15394.

157. Legault M, Perreault LL, Dubé M. ExPheWas: a browser for gene-based pheWAS associations medRxiv. 2021 doi: 10.1101/2021.03.17.21253824.

158. UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource . . 2007. Available: https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf.

159. Sheehan B. Assessment scales in dementia. Therapeutic advances in neurological disorders. 2012;5: 349-358 doi: 10.1177/1756285612455733.

160. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F *et al*. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. International journal of epidemiology. 2011;40: 1652-1666 doi: 10.1093/ije/dyr120.

161. UK Biobank. UK Biobank launches one of the largest scientific studies measuring circulating proteins, to better understand the link between genetics and human disease. . 2020. Available: https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/uk-biobank-launches-one-of-the-largest-scientific-studies.

162. Bretherick AD, Canela-Xandri O, Joshi PK, Clark DW, Rawlik K, Boutin TS *et al*. Linking protein to phenotype with Mendelian Randomization detects 38 proteins with causal roles in human diseases and traits. PLoS genetics. 2020;16: e1008785 doi: 10.1371/journal.pgen.1008785.

Appendix

All supplementary tables and figures are available at 10.5281/zenodo.5203043.

Supplementary Table 3.1

Odds ratios (OR) and p-values (P) when regressing sex on *KL* variants in the UK Biobank using logistic regression.

Supplementary Table 3.2

Genotype counts for Newcastle UK Biobank (UKB) and Newcastle 85+ (N85P) cohorts. CHISQ_2DF_P = p-value from a chi-squared test with 2 degrees of freedom. HOM = homozygous. HET = heterozygous.

Supplementary Table 4.1

Log odds ratios (beta) and their respective errors (SE) when regressing the six cardiometabolic outcomes on *KL* variants using logistic regression with and without adjusting for age and sex. N = sample size. A1 = effect allele.

Supplementary Table 4.2

Regression coefficients (beta) and their respective errors (SE) when regressing the 11 cardiometabolic traits on *KL* variants using linear regression with and without adjusting for age and sex. N = sample size. A1 = effect allele.

Supplementary Table 4.3

Regression coefficients (beta) and their respective errors (SE) when regressing HbA1c on 125 *KL* variants using linear regression adjusting for age, sex and rs576674. N = sample size. A1 = effect allele.

Supplementary Table 5.1

Standardised coefficients (beta), effect allele (A1), standard errors (se), sample sizes (n) and p-values (p) when regressing cognitive measures on 247 *KL* variants in the UK Biobank with (All) and without (≥69) including participants less than 69 years old, adjusted (adjusted) and unadjusted (unadjusted) for age, age^2, sex, and educational attainment.

Supplementary Table 5.2

Standardised coefficients (beta), effect allele (A1), standard errors (se), sample sizes (n) and p-values (p) when regressing cognitive measures on rs9536314 and on rs9527025 in the UK Biobank with a genotype*age interaction term and adjusted for age, age^2, sex, and educational attainment.

Supplementary Table 5.3

Standardised coefficients (beta), effect allele (A1), standard errors (se), sample sizes (n) and p-values (p) when regressing cognitive measures on rs9536314 and on rs9527025 in the UK Biobank in APOE4 carriers (E4) and non-carriers (EX), adjusted (adjusted) and unadjusted (unadjusted) for age, age^2, sex, and educational attainment.

Supplementary Table 5.4

Standardised coefficients (beta), effect allele (A1), standard errors (se), sample sizes (n) and p-values (p) when regressing cognitive measures on rs9536314 and on rs9527025 in the UK Biobank with a genotype*APOE4 interaction term and adjusted for age, age^2, sex, educational attainment, and APOE4 carrier status.

Supplementary Table 5.5

Standardised coefficients (beta), standard errors (se), sample sizes (n) and p-values (p) when regressing the rate of decline of cognitive measures on 247 *KL* variants in the UK Biobank, adjusted for age$_{T1}$, age$_{T1}^2$, sex, and educational attainment.

Supplementary Table 5.6

Standardised coefficients (beta), effect allele (A1), standard errors (se), sample sizes (n) and p-values (p) when regressing cognitive measures on rs141113969, on rs2227122, and on rs676046 in the UK Biobank with a genotype*age interaction term and adjusted for age, age^2, sex, and educational attainment.

Supplementary Figure 5.1

Standardised log odds (beta) and 95% Confidence Intervals when regressing Prospective Memory task performance on rs141113969 after ordering participants by age. Window size = 10,000. Step size = 1,000.

Supplementary Figure 5.2

Standardised log odds (beta) and 95% Confidence Intervals when regressing Prospective Memory task performance on rs141113969 after ordering participants by age and excluding those who performed the task at imaging. Window size = 50,000. Step size = 5,000.

Supplementary Table 6.1

ICD-10 codes used to define cases and controls for individual cancers and cancer categories.

Supplementary Table 6.2

Log odds ratios (beta) and their respective errors (se) when regressing cancers on *KL* variants using logistic regression with (adjusted) and without (unadjusted) adjusting for age and sex. N = sample size. A1 = effect allele. Model = genetic model used.

Supplementary Table 6.3

Hazard ratios (HR) and upper (upper) and lower (lower) 95% Confidence Intervals when regressing *KL* variants on cancer survival using a Cox proportional hazards model adjusted for age at diagnosis and sex. A1 = effect allele.

Supplementary File 7.1

A list file with one gene per row listing the gene name, the genomic region from which variants are chosen to be assigned to that gene and a list of variants that are assigned to that gene.