From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

DEMENTIA AND ITS COMORBIDITIES: GENETIC AND EPIGENETIC INFLUENCES

Ida K. Karlsson

Stockholm 2017

Cover illustration by Erik Sandkvist Nielsen Engebretsen.
Published by Karolinska Institutet.
Printed by Eprint AB.

ISBN 978-91-7676-767-2

Dementia and its comorbidities: Genetic and epigenetic influences

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Ida K. Karlsson

Principal Supervisor:
Professor Nancy L. Pedersen
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Co-supervisors:
Associate Professor Sara Hägg
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Dr. Alexander Ploner
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Opponent:
Professor Ruth Ottman
Columbia University
G.H. Sergievsky Center
College of Physicians and Surgeons

Examination Board:
Professor Agneta Nordberg
Karolinska Institutet
Department of Neurobiology, Care Sciences and Society (NVS)
Division of Translational Alzheimer Neurobiology

Associate Professor Lukas Käll
Kungliga Tekniska Högskolan
Division of Gene Technology

Associate Professor Fredrik Wiklund
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

“Sometimes I remember that I’ve forgotten. That’s the worst kind of forgetting.”

- And Every Morning the Way Home Gets Longer and Longer: A Novella. By Fredrik Backman.

ABSTRACT

Dementia is a multifactorial disorder of late life, characterized by memory deficits, personality changes, and impaired reasoning abilities. There is considerable co-morbidity between dementia, cardiovascular disease (CVD), and late-life depression, but the nature of the associations remains elusive. We therefore seek to investigate how genetic and epigenetic factors act, independently and in concert, to contribute to dementia as well as to its association with CVD and depression.

The first two studies focused on what role specific genes play in the association between dementia, depression, and CVD.

In **study I**, we investigated how apolipoprotein E (*APOE*) genotype affects the association between depression and dementia, and whether the timing of depression onset is of importance. Utilizing a nested case-control design with 804 dementia cases and 1,600 matched controls, we found that depression within ten years of dementia onset was associated with disease regardless of *APOE* genotype, while depression more distal to dementia was a risk factor only in carriers of the ε4 risk allele.

Study II focused on the shared genetic architecture between dementia and CVD, and entailed two parts. In the first part we used data from 13,231 Swedish twins, and found that genetically predisposed CVD was a stronger risk factor for dementia compared to CVD with a lower genetic risk. In the second part of the study we utilized summary statistics from previously published genome-wide association studies to investigate the genetic overlap between Alzheimer´s disease (AD), the most common form of dementia, and coronary artery disease. We found no evidence of genetic overlap between the disorders, but that both diseases have a significant number of genes in common with lipid levels.

The last two studies focused on epigenetic factors and investigated how gene specific methylation is associated with dementia.

Study III focused on the *APOE* gene, and how methylation levels in leukocytes relate to the risk of dementia, AD, and CVD. Using data from 447 Swedish twins, we demonstrated that hypermethylation in the promoter region of the gene was associated with dementia and AD, but not with CVD. Results were similar within discordant twin pairs, and did not differ as a function of *APOE* genotype.

In **study IV**, we focused on five other AD related genes that are differentially methylated in post-mortem brain samples from AD patients compared to controls. The aim was to investigate whether these differences could also be detected in blood samples collected pre-mortem. There was a significant difference in methylation of *SORL1* in leukocytes from dementia patients and of *BIN1* in leukocytes from AD patients. Findings were stronger in discordant twin pairs, indicating that the association cannot be attributed to genetic factors.

In conclusion, the studies included in this thesis highlight the complexity of late-life comorbidities, and the importance of taking both genetic factors and the timing of disease into account when studying these associations. Furthermore, methylation of genes related to AD is of importance for dementia, and has the potential to serve both as a biomarker and identify mechanisms of disease development.

LIST OF SCIENTIFIC PAPERS

I. **Ida K Karlsson**, Anna M Bennet, Alexander Ploner, Therese M-L Andersson, Chandra A Reynolds, Margaret Gatz, Nancy L Pedersen. Apolipoprotein E ε4 genotype and the temporal relationship between depression and dementia. *Neurobiology of Aging*, 2015, 36, 1751-1756; doi:10.1016/j.neurobiolaging.2015.01.008

II. **Ida K Karlsson**, Alexander Ploner, Ci Song, Margaret Gatz, Nancy L Pedersen, Sara Hägg. Genetic susceptibility to cardiovascular disease and risk of dementia. *Translational Psychiatry* (2017) 7, e1142; doi:10.1038/tp.2017.110

III. **Ida K Karlsson**, Alexander Ploner, Yunzhang Wang, Margaret Gatz, Nancy L Pedersen, Sara Hägg. Apolipoprotein E DNA methylation and late-life disease. *(Submitted)*

IV. **Ida K Karlsson**, Alexander Ploner, Yunzhang Wang, Margaret Gatz, Nancy L Pedersen, Sara Hägg. DNA methylation in Alzheimer´s disease associated genes. *(Manuscript)*

CONTENTS

LIST OF ABBREVIATIONS

Aβ	Amyloid-β
AD	Alzheimer´s disease
APOE	Apolipoprotein E
APP	Amyloid precursor protein
ATC	Anatomical Therapeutic Chemical
BDRS	Blessed Dementia Rating Scale
BMI	Body-mass index
CAD	Coronary artery disease
CDR	Causes of Death Register
CER	Cerebellum
CES-D	Center for Epidemiologic Studies Depression
CI	Confidence interval
CVD	Cardiovascular disease
DBP	Diastolic blood pressure
DSM	Diagnostic and Statistical Manual
DZ	Dizygotic
EWAS	Epigenome-wide association study
GRS	Genetic risk score
GWAS	Genome-wide association study
HDL	High density lipoprotein
HR	Hazard ratio
ICD	International Classification of Disease
IPT	In person testing phase
IRR	Incidence rate ratio
LDL	Low-density lipoprotein
meQTL	Methylation quantitative trait loci
MMSE	Mini-Mental State Examination
MZ	Monozygotic
NPR	National Patient Register
OR	Odds ratio
PDR	Prescribed Drug Register
PFC	Pre-frontal cortex
SBP	Systolic blood pressure
SD	Standard deviation
SNP	Single nucleotide polymorphism
STR	Swedish Twin Registry
T2D	Type 2 diabetes
TC	Total cholesterol
TG	Triglycerides
VaD	Vascular dementia
VEGAS	Versatile Gene-based Association Study

1 INTRODUCTION

While the global life-expectancy is steadily increasing, there are substantial differences in how we age. While some stay healthy throughout their lives others suffer from late-life disease and comorbidities from a relatively early age[1]. The determinants for these differences are largely unknown, and most age-related diseases are complex phenotypes with both genetic and environmental factors playing a role.

Dementia, which is the main focus of this thesis, is a debilitating late-life disorder with unclear mechanism and to date no effective treatment[2]. The incidence increases steeply with age, and as a result, prevalence is estimated increase exponentially due to the increasing life-expectancy[3]. If the trend continues, the number of individuals suffering from the disease is estimated to increase from around 47 million today to around 130 million in 2050[4]. Dementia is hence a major public health concern, and extensive work is needed to better understand disease mechanisms and develop a cure or disease-modifying therapy.

As the incidence of other late-life diseases also increases with age, co-morbidities are common in dementia. Diseases such as depression and cardiovascular disease (CVD) have been implicated as risk factors for dementia[5,6], but whether they are causal factors or simply associated due to shared risk factors remains to be elucidated.

Dementia is a complex multifactorial disorder, with both genetic and environmental factors influencing the risk of disease. Genetic factors explain 80 % of the variance in Alzheimer´s disease (AD)[7], the most common form of dementia, and several genes associated with the disease have been identified[8]. In recent years, epigenetic factors have also been studied extensively, and have been shown to be of importance for AD[9].

This thesis focuses on how genetic factors influence the association between dementia, depression, and CVD, and what role methylation of genes associated with AD play in disease development. The background will provide a short introduction to the field, and the methods used along with the main findings from each study will then be presented and discussed.

2 BACKGROUND

2.1 A short introduction to genetic epidemiology

2.1.1 Epidemiology and the problem of confounding

The field of epidemiology deals with the distribution of diseases and traits on the population level[10]. By studying the distribution of two traits in a population, usually an exposure and an outcome, we can learn something about their association. The basic idea is simple – if the exposed group more often experiences the outcome those exposed have an increased risk of the outcome compared to those unexposed.

However, reality is of course more complicated. The fact that two traits are associated at first glance does not mean there is a causal association. They could be associated through other factors connecting them – a concept known as confounding. One such example is retirement and dementia. At the adult population level it is certainly more common that dementia patients are retired than non-demented individuals, which of course does not mean that retiring causes dementia; rather, the association is heavily confounded by age, with those over a certain age being more likely to be retired as well as to have a higher risk of dementia. In addition, there is a wide range of systematic errors that can bias the apparent relationship between the exposure and outcome. Two common examples are selection bias, arising from studying a non-representative population, and information bias, arising from misclassification of the exposure and outcome. A special class of the latter is recall bias, where cases and controls may differ in how they report past exposures.

Even if all known confounders are adjusted for and other systematic errors dealt with, the nature of the observed association is still difficult to determine. First of all, an association does not imply causation. It may well be a result of residual confounding, factors you cannot control for or simply do not know about. It may also be a result of reverse causation, i.e. the outcome causes the exposure rather than the other way around. In studies of multifactorial phenotypes with unclear disease mechanisms, especially those with long progression periods, this is often a difficult problem to disentangle.

Classical epidemiological designs

Two commonly used designs in epidemiology are cohort studies and case-control studies.

In a cohort study, a group of individuals are followed over a certain time period, under which information about the outcome of interest is collected. The number of individuals who developed the outcome can then be compared between the exposed and unexposed participants. Collecting information in this prospective manner reduces the risk of recall bias and reverse causality as information about the exposure is collected prior to the outcome. The drawback is that cohort studies are often expensive and require a long follow-up time. Furthermore, if the outcome of interest is rare, very large sample sizes are required.

Case-control studies are, in contrast to cohort studies, retrospective in nature and normally based on a selection of cases combined with a representative selection of controls. The design is rather intuitive and easy to interpret; the distribution of the exposure among the cases is compared to that among the controls.

2.1.2 Genetic epidemiology

Genetic epidemiology is the study of genetic factors in health and disease, within families and across populations. It also deals with the interplay between genes and the environment and how it affects complex diseases and traits[11].

The extent of genetic and environmental influence on disease varies from Mendelian disorders caused by a single gene (such as Huntington's disease), to purely environmental factors (e.g. suffering a concussion after a head injury). Most diseases and traits fall in-between the two extremes, as complex multifactorial disorders are caused by a combination of many genetic and environmental factors acting in concert.

Twin studies

Twin studies have played a major role in determining the importance of genetic and environmental influences for a wide variety of traits. Twin studies build on the fact that monozygotic twins (MZ) are genetically identical while dizygotic (DZ) twins share on average 50% of their segregating genes. However, both types of twin pairs presumably share their environment to the same extent. Thus, the difference in correlation of traits within MZ versus DZ twin pairs can be used to estimate how much of the variance of a trait can be attributed to genetic and environmental factors[11].

The co-twin control design is another commonly used method in epidemiology, where the distribution of the exposure is compared within twin pairs discordant for the outcome of interest. They are a valuable contribution to the field, since some of the genetic and shared environmental factors are automatically adjusted for. Furthermore, the association between

exposure and outcome can be compared in DZ and MZ pairs to further elucidate the extent of genetic confounding.

GWAS and polygenic methods

Most identified genetic variants arise from mutations at a single site in the genome – a single nucleotide polymorphism (SNP). Over the past decade association studies have moved from focusing on candidate genes to hypothesis free genome-wide association studies (GWASs), analyzing SNPs throughout the entire genome. Thousands of SNPs associated with various phenotypes have been identified, and in many cases such findings have led to the discovery of new biological pathways involved in complex diseases[12]. Formation of GWAS consortia has led to substantial increases in the number of study participants, going from thousands up to hundreds of thousands. Despite the resulting increase in power, the genetic variants identified in most cases explain only a small proportion of the variance in a trait[13].

Several polygenic methods have been developed to utilize findings from GWASs to estimate heritability and investigate genetic architecture shared between phenotypes[13,14]. In addition to classical twin studies of heritability, it is now possible to estimate the variance in a trait explained by all the SNPs included on a genotyping array, using a sample of unrelated individuals. In most cases, the SNP based heritability is much lower than the heritability estimated from twin studies. The remaining part is called *missing heritability*, and may be explained by rare variants with large effect sizes, common variants with small effect sizes, variants not tagged by the arrays, non-additive genetic effects, and gene-environment interactions[13]. The effect of multiple SNPs on a trait can be summed up into a genetic risk score (GRS), which can be used as a measure of genetic susceptibility. The GRS can then be used to test how well the genetic susceptibility predicts the trait in different subgroups, or whether it predicts the risk of a different trait.

2.1.3 Epigenetics

Although all cells throughout the body carry the same genetic information, the genes expressed in different cell-types, and hence the cell functions, differ considerably. For example, the cells in the brain differ substantially in form and function from the cells in the skin, despite carrying the exact same DNA sequence. These differences are largely due to epigenetic factors, defined as mechanisms regulating gene expression through reversible mechanisms that do not alter the DNA sequence and are heritable through cell divisions[15]. These can hence modify the phenotype of a cell without modifying its genotype. Furthermore, epigenetic mechanisms have been proposed as a mechanism through which gene-environment interactions act[16].

Epigenetic mechanisms

Of the epigenetic mechanisms visualized in Figure 1, DNA methylation is the most extensively studied. It refers to a covalent modification where a methyl group is added to a cytosine located next to a guanine in the DNA, a so called CpG site[16]. Regions of DNA with unusually high CG content are referred to as CpG islands, and are often found in promoter regions of genes. Methylation in the promoter region inhibits the binding of transcription factors, and hence down regulates gene expression.

Two other epigenetic mechanisms are histone modification and noncoding RNA-mediated modulation of gene expression. In the chromosomes, DNA is carefully wrapped around nucleosomes with the help of histones[17]. By dynamic regulation of the histones, the density of the packed DNA can be modified. This determines the accessibility to transcription factors binding sites, and can thereby regulate transcription[15]. Non-coding RNAs are short RNA molecules that are not translated to proteins. The most well studied non-coding RNA are micro-RNA, which regulate gene expression at the transcriptional and post-transcriptional level by binding to mRNA[16].

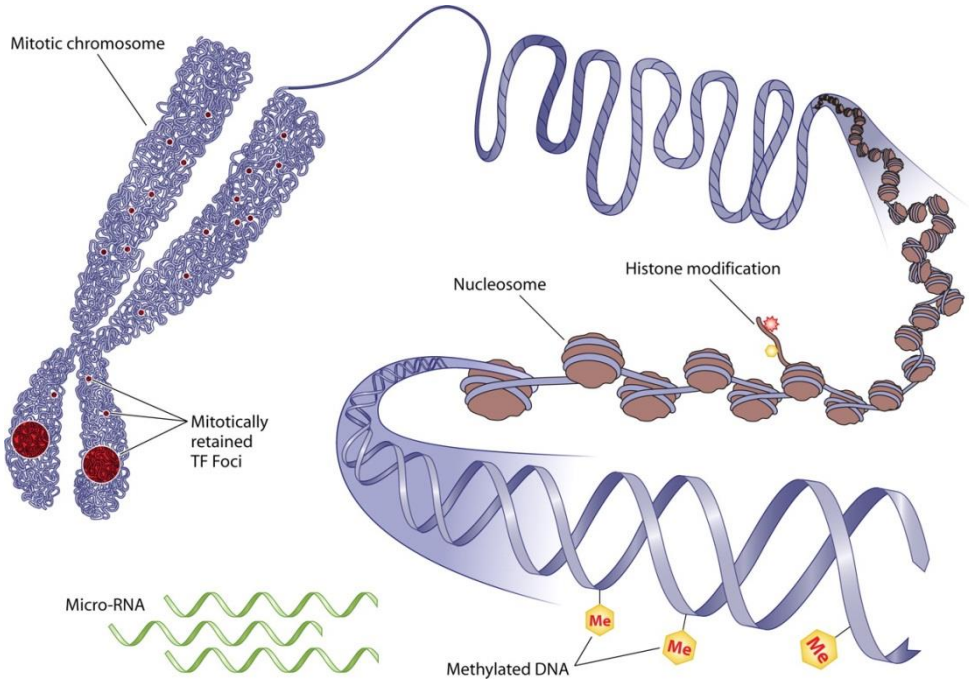


Figure 1: Epigenetic mechanisms. The figure depicts the most well studied epigenetic mechanisms: DNA methylation, histone modification, and noncoding RNA-mediated modulation of gene expression (Reprinted from Zaidi et al. 2010[15] with permission from Elsevier).

Epigenome wide association studies and interactions with genetic variants

Epigenome-wide association studies (EWASs) are based on the same hypothesis-free framework as GWASs, and investigate associations between a trait and methylation at specific CpG sites across the epigenome. At each CpG site, the proportion of methylated cells is analyzed for affected and unaffected individuals, which may indicate that either hypermethylation (higher levels of methylation) or hypomethylation (lower levels of methylation) is associated with the outcome of interest. Further investigations of the gene harboring the site, and whether it is located in the promoter region, may give further clues to the nature of the association.

However, associations between genetic variants and methylation complicate matters. Almost 20% of the variance in DNA methylation in blood is explained by genetic factors, and these genetic effects on methylation are stable throughout life[18]. Methylation quantitative trait loci (meQTLs) are genetic variants that influence methylation levels. While some meQTLs act in the surrounding region, the majority influence methylation elsewhere in the genome and identification of meQTLs affecting methylation levels at specific CpG sites can thus be challenging.

2.2 Dementia

Dementia is a multifactorial disorder with a long preclinical phase and poor understanding of underlying mechanisms. It is a geriatric disorder with the prevalence increasing steeply with age from around 1% in individuals aged 60-69 to approximately 30% in individuals aged 90 and older[19].

AD is the most common form of dementia, accounting for 60-80% of the cases, followed by vascular dementia (VaD), accounting for about 10% of cases[2]. However, evidence indicates that only a minority of dementia cases present either ‘pure AD’ or ‘pure VaD’ pathology, while most have a mixed pathology with elements of both AD and VaD[2,20].

2.2.1 Etiology

Alzheimer´s disease

AD exists in two forms, familial and sporadic AD. Familial AD has its onset before the age of 65, and is an autosomal dominant disease usually caused by mutations in the amyloid precursor protein (APP) or presenilin 1 or 2[21]. Sporadic AD on the other hand is a multifactorial complex disorder influenced by many genetic and environmental factors. The sporadic form accounts for more than 99% of cases[2] and is the focus of this thesis.

AD is a neurodegenerative disease characterized by accumulation of amyloid plaques and neurofibrillary tangles in the brain[5]. These changes begin as much as 20 years or more prior to symptom onset[2]. The main constituent of the amyloid plaques is amyloid-β (Aβ), which are formed from the cleavage products of the transmembrane protein APP. During cell metabolism, the extracellular tail of APP is cleaved either with α- or β-secretase. Cleavage by the latter enzyme results in Aβ, which is normally degraded or cleared from the brain. The amyloid cascade hypothesis is the most predominant hypothesis for the mechanism of AD development[5]. The hypothesis is based on an imbalance between the production and clearance of Aβ leading to aggregation and formation of neurotoxic amyloid plaques, thereby initiating AD pathology. APP and presenilin are key players in the amyloid pathway, and the fact that they are causative of the familial form of AD strongly supports the hypothesis.

Another hypothesis has the other pathological hallmark of AD, the neurofibrillary tangles, as central in disease development[5]. The neurofibrillary tangles are formed from hyperphosphorylated tau, a microtubule binding protein. Hyperphosphorylation of tau leads to dysfunction in microtubule assembly, causing impaired axonal transport and thus synaptic dysfunction. Hyperphosphorylated tau is also prone to aggregate into neurofibrillary tangles, further compromising neuronal function. It is still unclear whether this process is a cause or a consequence of AD[5].

Other hypotheses are based on the deficiency of acetylcholine, glutamate excitotoxicity, and neuroinflammation seen in AD[22]. In addition, dysfunction in blood vessels, oxidative stress, and mitochondrial dysfunction appear involved in the disease development, and it may well be that all these mechanisms are involved in the process to some extent[5].

Vascular dementia

VaD is a group of disorders characterized by vascular lesions disrupting the blood supply to the brain, thereby contributing to dementia development. Although there are familial cases, the sporadic form is most common[23]. The underlying mechanism is vessel disorders such as atherosclerosis of cerebral arteries or cerebral small vessel disease, leading to infarcts, white matter lesions, hemorrhages, or other types of vascular lesions[24]. These vascular lesions ultimately lead to disruption of the blood supply to affected areas of the brain and may hence cause impairment in brain function. Pure VaD without any elements of AD pathology is rare, and most often caused by infarcts[24]. It may arise either from several small infarcts in the cortical and subcortical regions of the brain (referred to as multi infarct dementia), or from single infarcts in regions important for cognition (strategic infarct dementia).

2.2.2 Diagnosis and treatment

Dementia criteria include significant cognitive impairment in one or more cognitive domains, severe enough to interfere with everyday living[25]. The diagnosis is based on medical and family history, neuropsychological tests, neurological examination, laboratory tests, and brain imaging. Dementia can then be further differentiated into subtypes, but autopsy studies are still considered the gold standard for a definitive subtype diagnosis[25,26].

To date, there is no treatment to slow down or stop the neurodegeneration, but there are approved pharmacological treatments for AD that can improve the symptoms temporarily[25]. These belong to two groups: cholinesterase inhibitors and memantine. Most therapeutic approaches target the acetylcholine deficiency (including the cholinesterase inhibitors), glutamate excitotoxicity (including memantine), clearance of Aβ, tau deposits, and neuroinflammation[22]. Despite large efforts and many new candidates developed, very few are successful. Between 2002 and 2014, 244 new drugs for AD were tested in clinical trials, out of which only one was successful[2]. Although many new compounds show promising results in Phase II clinical trials, adverse side effects or lack of therapeutic efficacy often leads to failure in Phase III[22]. Nevertheless, efforts are continuing and several drugs are currently undergoing Phase III trials.

2.2.3 Risk factors

Environmental risk factors

As mentioned previously, the most important risk factor for dementia is age, which has a great impact on disease risk. The disease is more common in women than in men, although the reason for the gender difference is not clear[2]. Low education is another well-established environmental risk factor for dementia, and a recent meta-analysis indicates that low educational attainment accounts for around 20% of the population-attributable risk[6]. Although the mechanism is not exactly clear it is presumed that education, as well as mental stimulation throughout life, maintains a cognitive reserve protecting against dementia[27]. Other risk factors associated with dementia are head injury, hypertension, high cholesterol, smoking, and obesity[5].

Co-morbidities

Co-morbidities are common in late life, and dementia is often seen together with other diseases such as vascular conditions, diabetes, cancer, and depression[5,6,28]. Cardiovascular disease is associated with dementia, but it is not clear whether CVD per se is a causal factor, or if common risk factors increase the risk of both diseases[5]. Physical inactivity, smoking, type 2 diabetes (T2D), and midlife hypertension and obesity are all vascular risk factors also associated with AD[6]. Several studies have shown that depression is associated with

dementia, but it is a complicated association where the age at depression onset appears to be of importance. A narrative review concluded that there is strong evidence that depression during midlife is a risk factor, while depression during late life is rather a prodromal stage of dementia[29]. In addition, the two diseases share several risk factors, and there is also an association between vascular disease and depression[30]. As evident, the associations between these late-life disorders are highly complicated, and the effects difficult to disentangle. Nevertheless, a better understanding of these associations opens up the possibility for implementing preventive strategies and treatment regimens that reduce the burden of more than one disease.

Genetic risk factors

Twin studies have estimated the heritability of AD at almost 80%[7], and a recent study estimated the SNP based heritability to be 53%[31]. The most important genetic risk factor is the apolipoprotein E (*APOE)* ε4 allele, which was identified in 1993[32,33] and has since been confirmed across a range of populations[21]. Heterozygote carriers have a threefold increased risk of AD, while the risk is as much as eight to twelve times higher in homozygote carriers compared to non-carriers[2]. The *APOE* ε2 allele decreases the risk of AD, but the effect is smaller than that of the ε4 allele. In addition to *APOE*, 19 loci associated with AD were identified in the most recent GWAS[8]. The genes implicated are involved in diverse mechanisms, such as APP and tau pathology, inflammatory processes and lipid transport. It should be noted that the effect sizes of these loci are not comparable to that of *APOE*. While *APOE* alone explains 13% of the total variance in AD, the other 19 known loci together explain 3%.[31].

Far less is known about the genetics of VaD. The only twin study performed to date showed the variation was due to environmental factors only, as no difference could be detected in concordance rates between MZ and DZ twins[34]. While candidate gene studies have identified genetic variants involved in lipid metabolism, inflammation, and angiotensin as associated with VaD[35], findings from GWASs have been limited[36,37].

Epigenetic risk factors

Some epigenetic studies have investigated the association between methylation, either gene-specific or epigenome-wide, and AD. A recent systematic review identified several studies that reported associations between methylation and AD using both blood and post-mortem brain samples[9]. Among the 38 studies included, the most consistent findings were in *BDNF*, *SORBS3*, and *APP*. The largest EWAS to date included 708 brain samples, and identified 11 differentially methylated sites[38], among others in *ABCA7* and *BIN1,* which also harbor genetic variants associated with AD[8]. Another of the identified sites resides in *ANK1*, which was also identified in a smaller EWAS study of AD published simultaneously[39].

Twin studies of methylation in dementia are rare. One study of AD discordant twins was based on a well-characterized monozygotic twin pair[40]. Both twins had the same education, but the affected twin was exposed to pesticides in his workplace while the other was not. The affected twin developed AD already at the age of 60, while the unaffected twin died from cancer at the age of 79. Using post-mortem brain samples, the authors identified significantly lower levels of global methylation in the affected AD twin than in the co-twin. A recent case study of another AD discordant twin pair examined methylation in the promoter region of selected genes using blood samples from the twins[41]. They found a general hypomethylation in both twins across all the promoters, but no difference between the affected and unaffected twin.

3 AIMS

The overarching aim of this thesis is to contribute to a better understanding of dementia etiology and the association with other late-life disorders, by studying the influence of genetic and epigenetic factors. The specific aims can be divided as follows:

Aim 1: Test how specific genes, individually and in combination, contribute to dementia and the association with depression and CVD (study I and II).

- Study I: Study how *APOE* genotype affects the association between depression and dementia, and whether the timing of depression onset is of importance.

- Study II: Investigate if genetic risk of CVD also increases the risk of dementia, and if it modifies the association between the two diseases; explore the shared genetic architecture between CVD, dementia, and their common risk factors.

Aim 2: Explore how DNA methylation in genes related to AD influence the risk of dementia (study III and IV).

- Study III: Evaluate whether DNA methylation of the *APOE* gene in leukocytes influences the risk of dementia and cardiovascular disease.

- Study IV: Test if differences in DNA methylation of AD related genes found in post-mortem brain samples can be replicated in leukocytes collected pre-mortem.

4 STUDY OVERVIEW

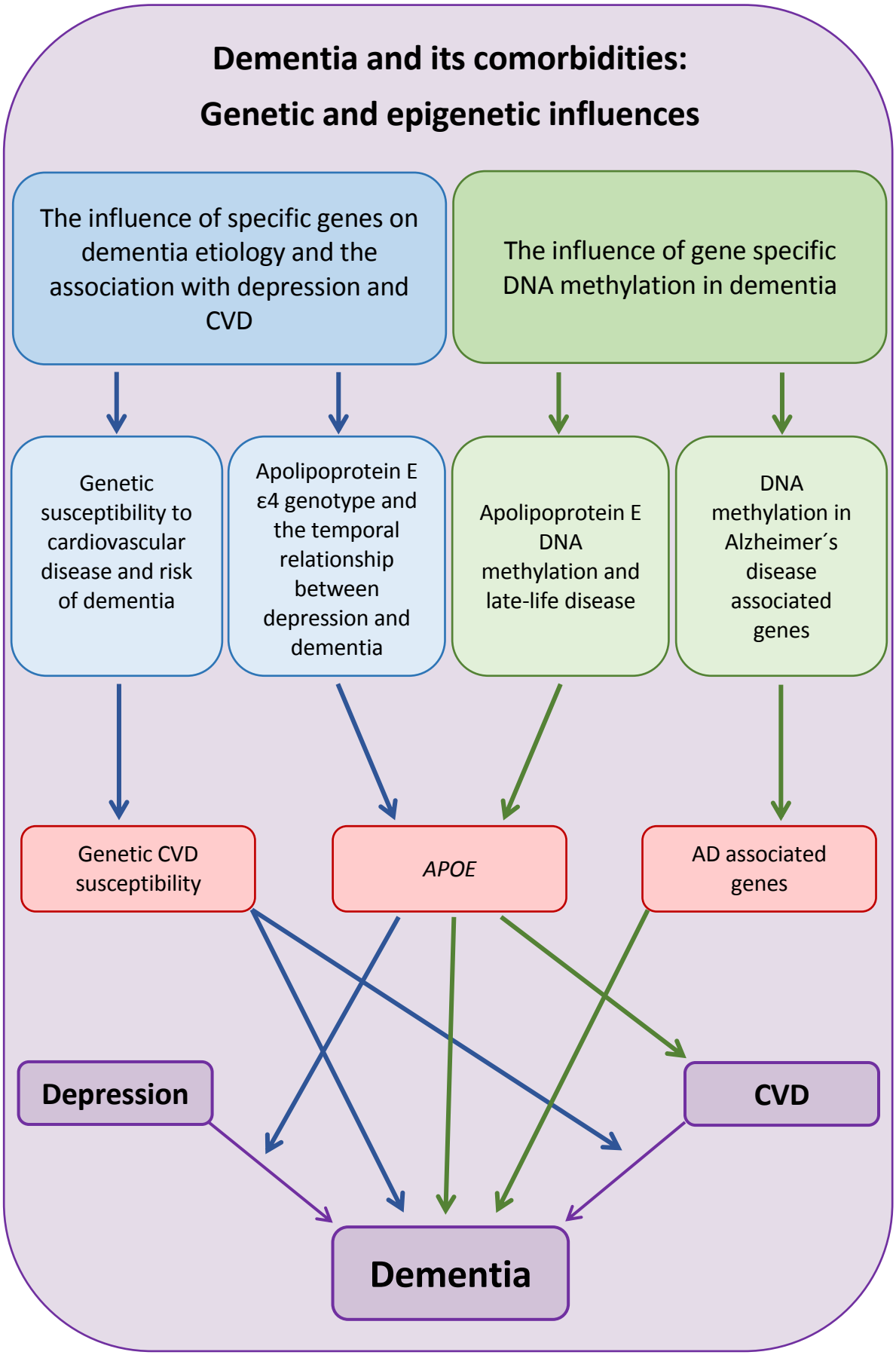


Figure 2: Overview of the four studies included in this thesis.

5 DATA

5.1 Data sources

5.1.1 The Swedish Twin Registry and sub-studies of aging

The Swedish Twin Registry (STR) was established in the 1950s, and is one of the largest twin resources in the world[42]. Information about all twin births was obtained from parish records or more recently, the National Board of Health and Welfare, and the registry currently includes 194,842 twins born between 1886 and 2008. Twins born prior to 1926 were included only if both members of the pair were still alive and had responded to a questionnaire, while all twins born from 1926 onwards were included, regardless of whether the twin partner was alive and had responded[19]. For 75,602 twin pairs, zygosity has been determined by an algorithm based on intra-pair similarity, DNA, or being of opposite sex. The accuracy of the intra-pair similarity algorithm has been shown to be 98%[42]. Within the STR, there are several sub-studies of aging. Figure 3 shows an overview of the five sub-studies included in this thesis.

SATSA

In the early STR questionnaires it was noted that a proportion of the twins had been separated during childhood. This was the basis of the longitudinal Swedish Adoption/Twin Study of Aging (SATSA)[43]. The study population includes all same-sex twin pairs from the STR who reported having been separated before the age of 11 and reared apart, matched to a sample of twins reared together based on sex, date and county of birth.

SATSA consists of both mailed questionnaires and in-person testing phases (IPTs). The first questionnaire (Q1) was sent out in 1984 to 2,845 individuals, out of whom 2,018 responded (71%)[43]. Questionnaires were then sent out every three years, except for a break between 1993 and 2004, to all individuals in the base-population, regardless of previous participation. The last questionnaire was sent out in 2010, and thus a total of 7 questionnaires were sent out as a part of the SATSA study.

The IPTs consisted of a health examination, cognitive tests, an interview, and collection of blood samples [43]. The first IPT was conducted between 1985 and 1988, and all twins over the age of 50, where both members of the pair answered Q1, were invited. In total, 645

individuals participated in IPT1. All twins who participated in the first IPT, together with those who answered Q1 and had turned 50 since IPT1, were invited to IPT2, and the same procedure was continued for IPT3. IPT4 was replaced with a telephone interview due to funding issues. From IPT5, the previous procedure was continued on a three-year rolling schedule through IPT8, and then on a two-year rolling schedule through IPT10, which was conducted 2012-2014 and marked the final phase of SATSA.

A total of 859 individuals from 449 same-sex twin pairs participated in at least one IPT, and are included in the study population used for this thesis. The mean age at first IPT participation was 63.6 years (standard deviation (SD) 8.8).

OCTO-Twin

Origins of Variance in the Old-Old: Octogenarian Twins (OCTO-Twin) is a longitudinal study of twins above the age of 80[44]. The study consists of IPTs on a two year rolling schedule, where the first wave was conducted between 1991 and 1994. Blood samples were mainly collected during the second wave. Same-sex twin pairs within the STR where both twins were still alive, at least 80 years of age during the first wave, and not already included in SATSA were eligible for the study. Out of the 549 pairs invited to the study, 351 participated in the first wave of testing (pairwise response rate 64%, corresponding individual response rate 80%)[45]. The mean age at the first wave was 83.6 years (SD 3.2).

GENDER

The Aging in Women and Men (GENDER) study is a longitudinal study of opposite-sex twin pairs born between 1906 and 1925[46]. The study includes three IPTs on a four year rolling schedule, similar in content to the IPTs in SATSA. Blood samples were collected as part of the first IPT. All opposite-sex twin pairs where both members were still alive were identified in the STR and sent a questionnaire (n=1,699 pairs). In total, 1,843 individuals (54%) replied, leading to 602 complete pairs (pairwise response rate 35%). All pairs aged 70 to 79 between 1995 and 1997 were invited to the first IPT, in which 249 pairs participated. Mean age at baseline was 74.6 years (SD 2.6).

HARMONY

The Study of Dementia in Swedish Twins (HARMONY) study is a cross-sectional study of both same-sex and opposite-sex twin pairs[19]. It is based on the Screening Across the Lifespan Twin (SALT) study, a telephone interview aimed at all twins born 1958 or earlier conducted between 1998 and 2003[47,48]. For all twins aged 65 or older, a tool aimed to identify dementia cases was added to the SALT interview (n=20,269). This sub-sample is known as the HARMONY sample. A total of 14,435 (71%) individuals participated (in 712 cases, an informant was interviewed), out of which both members of 4,537 pairs were included[19].

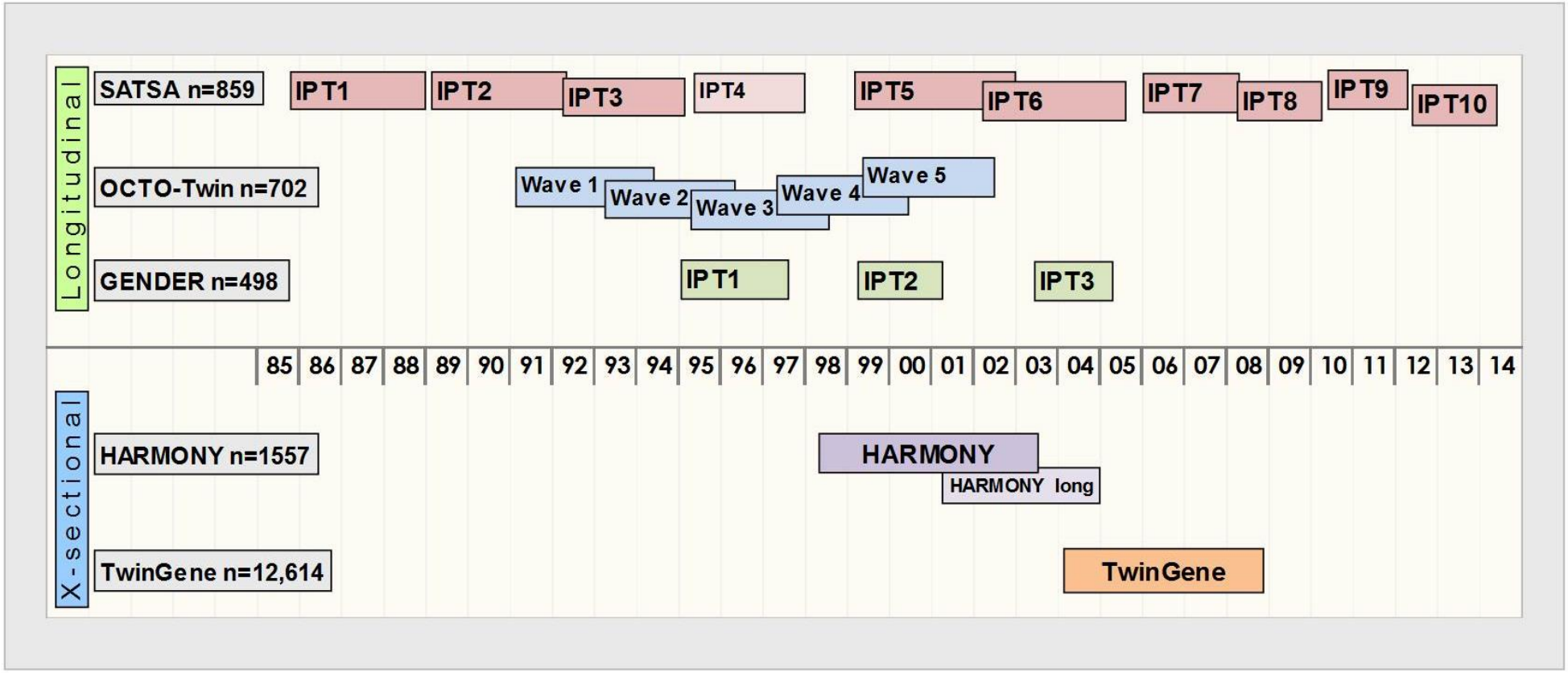


Figure 3: Overview of the five sub-studies of aging within the Swedish Twin Registry. The name of each sub-study together with the number of participants are given in the first box. The colored boxes indicate the different testing phases. IPT=in-person testing phase. Modified from Eriksson 2010[49].

Mean age was 73.5 years (SD 6.6). All individuals who were suspected of having dementia were referred for a clinical workup that included physical and neurological examination, an extensive interview, and collection of blood samples. If the preliminary assessment indicated the twin was demented, their twin partner, regardless of screening status, was also referred for workup. In addition, a control sample of 35 twin pairs where both members screened negative were also referred for the clinical workup. Out of 2,139 individuals invited to the workup, 1,557 participated (overall participation rate 73%). In addition, for 156 individuals who were deceased co-twins of HARMONY participants, dementia diagnoses based on medical records were included[19]. Twins who participated in the clinical workup are included in this thesis.

A subsample of twins diagnosed as questionable dementia, and non-demented partners of demented twins with disease onset within five years prior to assessment were invited to a longitudinal follow-up phase.

TwinGene

TwinGene is a study of twins born between 1911 and 1958 where both members were still alive and had participated in the SALT study[42]. TwinGene was conducted between 2004 and 2008. It entailed a questionnaire with questions about common diseases and a health checkup where blood samples were collected. In total, 22,390 twins were invited to participate in the study. Out of those, 12,614 individuals, including 5,014 complete pairs, gave their consent and left a blood sample. Hence, the individual response rate was 56% and the pairwise response rate 45%. The mean age was 64.9 years (SD 8.1).

5.1.2 Genetic and epigenetic data

Genetic information

Genotype information was available for 13,258 individuals. Among the TwinGene participants, 10,714 were successfully genotyped using Illumina Human OmniExpress, and imputed against 1000 Genomes Project phase 1 version 3 data[50]. Among the participants in SATSA, OCTO-Twin, GENDER, and HARMONY, 2,702 were genotyped using CardioMetabochip[51]. This is a customized chip where loci of importance to cardiovascular and metabolic traits are prioritized. For 158 twins, genotype information was available from both sources. In most cases, only one member of MZ twin pairs was genotyped, and the co-twin's genotype imputed.

The *APOE* ε2 and ε4 SNPs were not available on the CardioMetabochip, and were therefore directly genotyped in 2,999 twins from SATSA, OCTO-Twin, GENDER, and HARMONY. For TwinGene participants, *APOE* ε2 genotype was included on the chip, and ε4 imputed based on the 1000 genomes panel according to a protocol with high accuracy[52].

Epigenetic information

Methylation information from blood samples was available for 62 twins from HARMONY and 385 twins from SATSA. In the latter, up to five measurements per individual were available (blood samples collected during IPT3, IPT5, IPT6, IPT8, and IPT9), and a total of 1094 samples were available from 447 individuals.

In addition, methylation information from post-mortem brain samples was available from 39 individuals from SATSA and HARMONY, from cells in both the pre-frontal cortex (PFC) and cerebellum (CER). Out of those, 29 are included among the 447 individuals with methylation information from blood samples.

Methylation levels in leukocytes were analyzed with the Infinium Human Methylation 450K BeadChip (Illumina Inc., San Diego, CA, USA), and in neuronal cells with the Infinium MethylationEPIC BeadChip (Illumina Inc., San Diego, CA, USA). DNA was first extracted and bisulfite converted using the EZ-96 DNA MagPrep methylation kit for leukocytes and the EZ DNA Methylation Gold Kit for neuronal cells (both from Zymo Research Corp., Orange, CA, USA), and then hybridized to the bead chips. The obtained methylation data were pre-processed using a rigorous multi-step quality control pipeline. Samples were removed if they showed poor correlation to genotype controls or had the wrong sex predicted. Probes were removed if they had detection p-value above 0.05, overlapped with a SNP site, or resided on sex chromosomes. Processing was performed in R with background correction done using methylumi.noob[53], and normalization using wateRmelon.dasen[54]. For the leukocyte samples, the Houseman method[55] was used to adjust for cell counts, and sva::Combat[56] to adjust for batch effects. Due to the low number of neuronal samples, it was not possible to adjust for batch effects, and slide ID was therefore added as a covariate in all analyses of neuronal cell methylation.

The obtained beta-values are the ratio of methylated to total (methylated plus unmethylated) probe intensity for each CpG site, which in essence can be interpreted as the percentage of methylated cells[57]. Although easy to interpret, the beta-value has poor statistical properties, with heteroscedasticity in the lower and higher range, violating the assumption of many statistical models. Therefore, the beta-values were logit2-transformed into M-values. Although the M-values do not have a direct biological interpretation, they have far better statistical properties[57] and were therefore used throughout this thesis.

5.1.3 National healthcare registers

Since 1947, all Swedish residents are assigned a personal identification number, consisting of a six digit birthdate plus a four digit identification number[58]. The personal identification number is used for all public administration, including healthcare, and can thus be used for linkage across registers. The STR is linked to several population-based registers, of which the

National Patient Register (NPR), the Causes of Death Register (CDR), and the Prescribed Drug Register (PDR) are used here. Register information is obtained for all twins who have actively participated in the STR, meaning that they must have participated in one of the sub-studies or replied to at least one questionnaire. The registers are administered by the National Board of Health and Welfare (Socialstyrelsen in Swedish). All Swedish residents are covered by universal health insurance covering most of the cost for healthcare and medications, with only a small fee paid by the patients.

National Patient Register

The aim of the NPR is to follow the general health of the population, facilitate prevention and treatment of disease, and contribute to healthcare development[59,60]. The register was initiated in the 1960s, then covering 6 of the 26 counties in Sweden at the time. The register expanded, and by 1983 covered 85% of all hospitalizations. In 1984, it was decided that participation should be mandatory for all counties, and the NPR reached nationwide coverage of all overnight hospitalizations in 1987. Since 2001, the register also includes all outpatient specialist care, from both public and private caregivers. Primary care is currently not covered by the NPR.

Each record corresponds to one hospitalization or special care visit, with the primary diagnosis and additional diagnosis classified according to the International Classification of Diseases (ICD) codes. In total, 99% of the inpatient records and 80% of the outpatient records have information about the diagnosis. The NPR is updated yearly, and the most recent linkage to the STR includes information through 2014.

Causes of Death Register

Through the CDR, the causes of death and subgroup specific mortality in Sweden can be followed and evaluated[61,62]. The register was established with nationwide coverage in 1961, and contains information about the underlying as well as contributing causes of death reported in ICD codes. Up to 2011, the CDR includes information about the death of all individuals who in the year of their death were registered as residents in Sweden, regardless of whether the death occurred in Sweden or abroad. From 2012, it includes information about all deaths occurring in Sweden, regardless of residency, as well as about the deaths of Swedish residents occurring abroad. Like the NPR, the CDR is updated yearly, and the linkage to the STR contains information on causes of death through 2014.

Prescribed Drug Register

The PDR was initiated 2005, with the aim of increasing patient safety regarding medications[63,64]. It includes information about all dispensed medications, reported according to Anatomical Therapeutic Chemical (ATC) codes. It does not include information

about over-the-counter medications, medications used in hospitals or care facilities, or medications prescribed but not dispensed. In addition to information on the type of medication, the register also includes information about the prescribed dose, date of prescription, and date of dispense. The register is updated monthly, and the latest linkage to STR includes information through 2015.

5.2 Disease ascertainment

5.2.1 Dementia ascertainment

We used two sources to obtain information about dementia, AD, and VaD. Dementia was clinically evaluated as part of the SATSA, OCTO-Twin, GENDER, and HARMONY studies. In addition, dementia information was available from the nation-wide health registers previously mentioned.

Clinical ascertainment

In SATSA, OCTO-Twin, and GENDER, the Swedish version of the Mini-Mental State Examination (MMSE) [65] was used as a screening tool for dementia. The face-to-face interview requires only 5-10 minutes to administer, and includes 11 questions focused on the cognitive aspects of mental function. The MMSE is extensively used to screen for dementia, both in a clinical setting and in research. The score correlates with age and education, and cut-offs can be adjusted to fit the study population[66]. The maximum score is 30, and scores of 24 or higher were considered normal, while lower scores indicated mild (19 to 23), moderate (13 to 18), or severe (12 or less) cognitive dysfunction.

In HARMONY, the TELE[67], a telephone assessment for screening of dementia, was used. The TELE is based on the 10-item Mental Status Questionnaire (MSQ), supplemented with other cognitive items and questions about health and daily functioning. For those who performed poorly, an informant was interviewed using the Blessed Dementia Rating Scale (BDRS)[68] developed to assess the ability to deal with practical tasks of everyday life. The TELE and BDRS scores were then combined and transformed into an ordinal scale ranging from zero to three, with zero being cognitively intact and three indicating cognitive dysfunction.

In HARMONY and up to the third IPT of SATSA, all twins who screened suspect for dementia and their co-twins were referred to a clinical evaluation. In large, the workup followed the Consortium to Establish a Registry for Alzheimer´s Disease (CERAD) protocol[69], including physical and neurological examinations, informant interviews, review of medical records, neuroimaging, and laboratory tests. In OCTO-Twin and GENDER, and from the fifth IPT of

SATSA, dementia diagnoses were based on the cognitive tests included in the IPTs, review of medical records, and the research nurses evaluation of the twins.

For all four sub-studies, final dementia diagnoses were set at multidisciplinary consensus conferences, following Diagnostic and statistical manual of mental disorders (DSM)-III-R[70] or DSM-IV[71] criteria. Differential diagnosis of AD followed NINCDS/ADRDA criteria[72], and that of VaD followed the NINDS/AIREN criteria[73] .

Dementia from registers

Information about dementia diagnoses was extracted from the NPR, using both the primary and additional diagnoses. The date of admittance (for inpatients) or diagnosis (outpatients) was used as dementia onset. Dementia as the primary or contributing cause of death was extracted from the CDR for deceased individuals. For cases identified only through the CDR, we estimated age at onset by subtracting the mean number of years between the first dementia diagnosis in NPR and a record of dementia in CDR for individuals who had both. ICD codes used to identify dementia cases are shown in Table 1.

In addition, information on dispensed dementia medication was retrieved from the PDR and used as a proxy to identify cases. All dementia medication currently prescribed in Sweden is specific to AD, and all cases identified through the PDR were categorized as such. All medications in the ATC group N06D were included (namely N06DA02-04 and N06DX01). Age at first record of prescription was used as age of disease onset.

5.2.2 Depression ascertainment

Depression information was available from four sources, namely the NPR, review of medical records, antidepressant use, and the Center for Epidemiologic Studies Depression (CES-D) scale.

Information about depression from the NPR was available for all participants, and the ICD codes used are presented in Table 1. Medical records were reviewed as part of the SATSA, OCTO-Twin, GENDER, and HARMONY studies, and information about depression diagnosis as well as use of antidepressant medication was collected. In addition, self-reported use of antidepressant medication was available from the questionnaires in SATSA, OCTO-Twin, and GENDER. The CES-D scale was administered during every IPT in the SATSA, OCTO-Twin, and GENDER studies. The CES-D scale is a 20-item scale of current depressive symptoms, developed for epidemiological studies[74]. A score of 16 is normally used as a cut-off to identify depression cases, but in the this sample we used 20 as the cut-off, as this has been shown to better identify depression cases among elderly individuals[75]. Information from all four sources was combined, and the first record of depression was used as age at onset.

Table 1: ICD codes used to identify dementia and CVD cases

	ICD-7	ICD-8	ICD-9	ICD-10	Surgical code
Dementia					
Alzheimer´s disease	304-305	290	290A/B	F00	
			331A	G30	
Vascular dementia		293.0-1	290E	F01	
Other dementia	306		290X/W	F02-03	
			294B	G311	
			331B/C/X	G318A	
				F051	
Depression					
	314.99	296.00	296C/D/W	F32-33	
		298.00	298A	F34.1	
		300.40-41	300E	F41.2	
		790.20	309A/B		
			311X		
CVD					
Non-stroke CVD	420	410-414	410-414	I20-25	984
	450	440	440	I79	3068
	453.33	443.90	443X	I73.9	3080
					3127
					3141
					3158
					FNC/FND/FNE
					FNG00/02/05
CAD		410-411	410	I20.0	3080
			411B	I21-22	3127
					3158
					FNC/FND/FNE
					FNG02/05
Stroke	330	430-431	430431	I60-61	
	331.00-01	433-434	434	I63-64	
	331.09/99	436	436		
	332.00-19				
	332.29				
	334.00-98				

5.2.3 Cardiovascular disease ascertainment

Information about CVD was available from the NPR and the CDR. The ICD codes used are presented in Table 1, and are separated into stroke and non-stroke CVD. Stroke includes both ischemic and hemorrhagic stroke. Non-stroke CVD includes atherosclerosis, claudication, unstable angina, myocardial infarction and the surgical procedures coronary artery bypass grafting and percutaneous transluminal coronary angioplasty. The date of admittance (for inpatient), diagnosis (outpatient), or death (CDR) was used to determine age at onset. A more strict definition of coronary artery disease (CAD) was also used where only primary diagnoses of myocardial infarction or unstable angina were included.

6 STUDY DESIGNS

6.1 Study I: *APOE* ε4 genotype and the temporal relationship between depression and dementia

In study I we investigated how *APOE* genotype influences the association between depression and dementia, while also taking the timing of depression into account.

The study is based on the SATSA, OCTO-Twin, GENDER, and HARMONY studies, using dementia information from the clinical ascertainment (available for 2,884 individuals). It utilized a nested case-control design, which is a variant of the classical case-control design, but also has some of the advantages of the cohort design. For each dementia case we randomly selected two controls matched on sub-study, sex, and year of birth within two years. The controls had to be cognitively intact and still participating in the study at the age of dementia onset in the case. Controls may, however, develop the disease themselves later on. Hence, a case may be included more than once, both as a case and as a control for another case. Similarly, the same control may be randomly selected for more than one case. This is referred to as incidence density sampling[76], and allows for a representative sample of the exposure status and person-time at risk to be obtained. To avoid matching cases to their co-twins or to two controls from the same twin pair, only one member of a pair was allowed in each matching stratum. The selection resulted in 804 dementia cases and 1,600 matched controls, and included 1,519 unique individuals.

If late-life depression is indeed a prodromal stage of dementia, it is plausible that time from depression onset to dementia diagnosis better captures this than the age at depression onset. Hence, to better understand the importance of the timing of depression, it was categorized in two ways. First in relation to time to dementia diagnosis, with one category having a first record of depression onset within ten years of dementia, and one having their first depression onset more than ten years prior to dementia diagnosis. Secondly, we categorized depression according to age at onset, with late-life depression having a first record of onset at the age of 60 or later and midlife depression if the first record was before the age of 60.

APOE genotype was categorized into carriers (genotype ε3/ ε4 and ε4/ ε4) and non-carriers (all other genotypes) of the ε4 allele. Individuals with the ε2/ ε4 genotype were considered non-carriers due to the protective effect of the ε2 allele.

We performed conditional logistic regression using SAS 9.3. Covariates included, in addition to the matching variables age, sex, and sub-study, were education (dichotomized into seven years or less, versus more than seven years), stroke prior to dementia onset in the case, and source of depression information. We first modeled the effect of depression (using both types of categorization) on dementia and AD. Secondly, we introduced an interaction term between depression and *APOE*, and modeled the effect of depression on dementia and AD stratified on *APOE* genotype.

6.2 Study II: Genetic susceptibility to CVD and risk of dementia

Study II consisted of two parts. The first part focused on the association between non-stroke CVD and dementia, and how genetic susceptibility to CAD influences the risk of dementia and the association between the two diseases. In the second part, we utilized summary statistics from published GWASs to investigate the genetic overlap between CAD and AD, as well as with their shared risk factors.

In the first part, we included all 13,231 participants in the sub-studies who had genotypic and dementia information available. We utilized a cohort design, following individuals from 1978 or the age of 50, until the end of 2014, death, or dementia onset, whichever occurred first. All available dementia information was included, and in addition to dementia, AD and VaD were modeled as separate outcomes. Non-stroke CVD events were retrieved from the NPR and modeled as a time-dependent exposure. Individuals were considered unexposed up until the time of CVD diagnosis, followed by one exposure level during the first three years after diagnosis, and a second exposure level more than three years after diagnosis[77]. A GRS for CAD was created based on the CARDIoGRAMplusC4D consortium's most recent GWAS, where they identified 55 SNPs associated with CAD [78]. For each individual, the number of risk alleles at each locus was summed up to an un-weighted score, which was used as a measure of genetic susceptibility to CAD. All models were adjusted for age, sex, education (dichotomized into more or less than 7 years), and T2D. Relatedness among the twins was accounted for by using robust sandwich estimators.

The Cox proportional hazard model with age as the underlying timescale was used in three main models: 1) the effect of the CAD GRS on dementia, 2) the effect of non-stroke CVD on dementia, and 3) the effect of non-stroke CVD on dementia, stratified on quartiles of the CAD GRS. All analyses were performed using STATA 13.

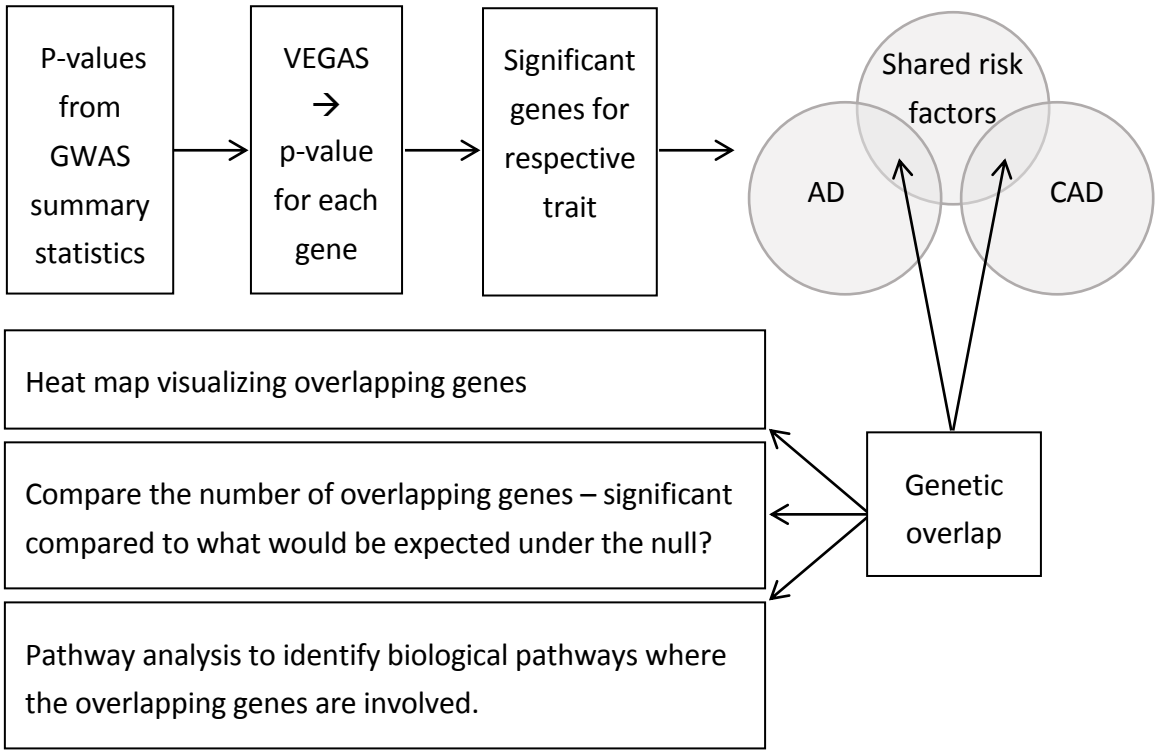


Figure 4: Workflow in the second part of study II. For each phenotype, SNP p-values from GWAS summary statistics were converted to gene-based p-values using the VEGAS approach. Significant genes for each outcome were then compared, and genes overlapping between AD or CAD and at least one of the shared risk factors were visualized in a heat map. In addition, we tested the statistical significance of the number of overlapping genes. We also used pathway analysis to identify biological pathways in which the overlapping genes are involved. GWAS, genome-wide association study; Vegas, Versatile Gene-based Association Study; AD, Alzheimer´s disease; CAD, coronary artery disease.

In the second part of study II, we utilized summary statistic from GWASs of AD[8] and CAD[78], as well as their shared risk factors body-mass index (BMI)[79], T2D[80], systolic- and diastolic blood pressure (SBP and DBP)[81] and the lipid fractions high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides (TG), and total cholesterol (TC)[82]. The work-flow is summarized in figure 4. For each set of summary statistics, we used the 'Versatile Gene-based Association Study' (VEGAS) approach[83] to obtain gene-level p-values. The method combines p-values from all SNPs within a gene into a gene-based p-value, while also accounting for the correlations among the included SNPs. In total, 17,581 genes were included, and genes with p-value below 2.84×10^{-6} were considered significant after Bonferroni correction. Genes were considered as overlapping if they were significant for more than one phenotype. A heat map was created visualizing the number of significant genes shared by either AD or CAD and at least one of the other phenotypes. The number of overlapping genes between two traits was compared to what would be expected under the null hypothesis, and the significance of the overlap was calculated using a binomial test. To

identify biological pathways in which the overlapping genes are involved, we used the Consensus Path Database[84]; this web-based tool allows testing for pathway enrichment of selected genes, adjusting for multiplicity by presenting false discovery rate corrected p-values as q-values.

6.3 Study III: *APOE* DNA methylation and late-life disease

In study III, we investigated leukocyte DNA methylation in the *APOE* gene in relation to dementia and CVD.

We used information from the last available blood sample for each of the 447 individuals with methylation information. Dementia, AD, and CVD were modeled as separate outcomes. Dementia information from both the clinical ascertainment and from registers was used to identify cases. CVD cases were identified from the NPR and CDR, and included diagnoses of both stroke and non-stroke CVD. The CpG sites within the *APOE* gene were categorized into three regions based on previous work[85]. Region 1 included CpG sites in the promoter region, region 2 sites residing in the first two exons and introns, and region 3 covered the fourth exon, which also harbors the ε2 and ε4 alleles. The mean methylation level across the sites in each region was used in the models.

All analyses were performed in STATA 13. We first used unconditional logistic regression models to test the association between methylation levels in each region and the three outcomes. Covariates included were age at blood draw, sex, and smoking. Relatedness of the twins was accounted for by using robust sandwich estimators. Secondly, we used conditional logistic regression to compare methylation levels in discordant twin pairs. These models were adjusted for age at blood draw and smoking.

To study genotype specific effects, we included an interaction term between *APOE* genotype and the CpG regions in the unconditional logistic regression model described above. *APOE* genotype was categorized into ε3/ε3, ε2 carriers (ε2/ε2 and ε2/ε3), and ε4 carriers (ε3/ε4 and ε4/ε4). While the ε2 allele decreases the risk of dementia and the ε4 allele increases the risk, both alleles increase the risk of CVD. Individuals with ε2/ε4 were therefore excluded from the analyses in order to investigate the effect of each of the alleles individually.

To compare the association between cases already diagnosed at blood sample to those with incident disease, dementia, AD, and CVD were divided into prevalent and incident cases and modeled separately in the unconditional logistic regression model.

6.4 Study IV: DNA methylation in AD associated genes

A previous study by Yu et al. investigated methylation levels in 28 AD related genes in post-mortem brain samples from the PFC[86]. They found five of the genes to be differentially methylated in AD cases compared to controls, namely *SORL1, ABCA7, HLA-DRB5, SLC24A4*, and *BIN1*. The primary aim of study IV was to try to replicate these findings in leukocytes collected prior to death. In addition, we also aimed to replicate the findings in post-mortem brain samples, and to investigate how methylation in leukocytes for those genes is correlated with methylation in neuronal cells.

For this purpose, we utilized the methylation data from leukocytes as well as from PFC and CER samples. From the leukocyte samples, we excluded individuals selected for Parkinson's disease (n=16) or where the blood sample was collected prior to the age of 60 (n=4). For the remaining 427 individuals we selected the last available sample prior to disease onset for dementia cases, alternatively the first sample available after disease onset. To obtain a similar age at blood draw, we selected the last available sample for controls. All 39 samples from neuronal cells were included.

Dementia from the clinical ascertainment combined with information from registers after end of follow-up was used as the outcome for analyses of leukocyte methylation. As part of AD diagnosis during autopsy, the amount and spread of neurofibrillary tangles in the brain are measured and categorized into Braak stage 0-V[87]. For a better measure of AD pathology at the time of death, we used Braak staging as the outcome for analyses of neuronal cell methylation. For all three tissue types, we selected all available CpG sites within and 100Kb around the five genes of interest.

In the first step, we ran regression models to study the association between each CpG site and dementia, AD, or Braak stage. For the leukocyte samples, we used a logistic regression model adjusted for sex and age at blood sample or death, with robust sandwich estimators to correct for relatedness. In addition, we ran conditional logistic regression to study the differences in methylation within discordant twin pairs. The co-twin control models were adjusted for age at blood sample. For the neuronal cell samples, we used linear regression of methylation in relation to Braak-staging. These models were adjusted for sex, age at death, and batch effects.

In the second step, we used the Fisher product method to combine the p-values from the regression models into a test statistic for each gene. The statistical significance of the obtained test statistics was tested using random permutations (n=1,000) with α=0.05 as the significance level. The null hypothesis underlying this method is that none of the CpGs in the gene are associated with the outcome.

To investigate whether methylation levels correlate in leukocytes, PFC, and CER cells, we calculated Pearson correlations for methylation levels in each CpG within the genes across the three tissues. All analyses were performed in R 3.3.2.

7 MAIN RESULTS AND INTERPRETATION

7.1 Study I: *APOE* modifies the association between depression and dementia, but only for depression more than ten years prior to dementia onset

Out of the 804 dementia cases, 469 were diagnosed with AD, and the mean age at dementia onset was 78 years. A total of 424 out of the 1519 unique individuals included in the study met the criteria for depression. Of these, 332 had their first record of depression within ten years of dementia, and 92 more than ten years prior to dementia onset in the case. The mean age at onset for depression was 74 years. The age range of depression within ten years of dementia onset was 58-90 years, and that of depression more than ten years prior to dementia was 23-79 years.

The association between depression and dementia was stronger for depression within ten years of dementia onset (incidence rate ratio (IRR) 4.46, 95% confidence interval (CI) 3.44-5.76 for dementia, IRR 3.45, 95% CI 2.39-4.98 for AD) than for depression with onset more distal to dementia (IRR 1.58, 95% CI 1.07-2.34 for dementia, IRR 1.75, 95% CI 1.01-3.03 for AD). When stratifying on *APOE* ε4 genotype, the association between depression within ten years of dementia and disease was similar in the two genotype categories (Figure 5). However, depression with onset more than ten years prior to dementia onset significantly interacted with *APOE* ε4 genotype (p=0.01 for both dementia and AD) and was a risk factor only in carriers of the ε4 allele.

The same overall pattern was seen when depression was categorized into late-life versus midlife depression, but interestingly the difference between the two depression categories was not as evident (late-life depression: IRR 3.56, 95% CI 2.81-4.51 for dementia, IRR 2.93 95% CI 2.08-4.12 for AD. Midlife depression: IRR 2.43, 95% CI 1.35-4.35 for dementia, IRR 2.63, 95% CI 1.13-6.09 for AD).

While the association between late-life depression and dementia has been relatively robust in previous studies, the association between depression earlier in life and dementia has been more inconsistent[88]. Our findings may shed light on the complexity of the association in two ways. Firstly, the effect of depression within ten years of dementia was robust regardless of *APOE* genotype, while depression more distal to dementia was a risk factor

only in carriers of the ε4 allele. Importantly, this shows there are differences between depression occurring close in time to dementia onset and depression occurring earlier in life, further strengthening the theory of depression as a prodromal feature. This also indicates that the discrepancies in previous studies of depression earlier in life and dementia may to some extent be explained by genetic factors, such as *APOE* genotype.

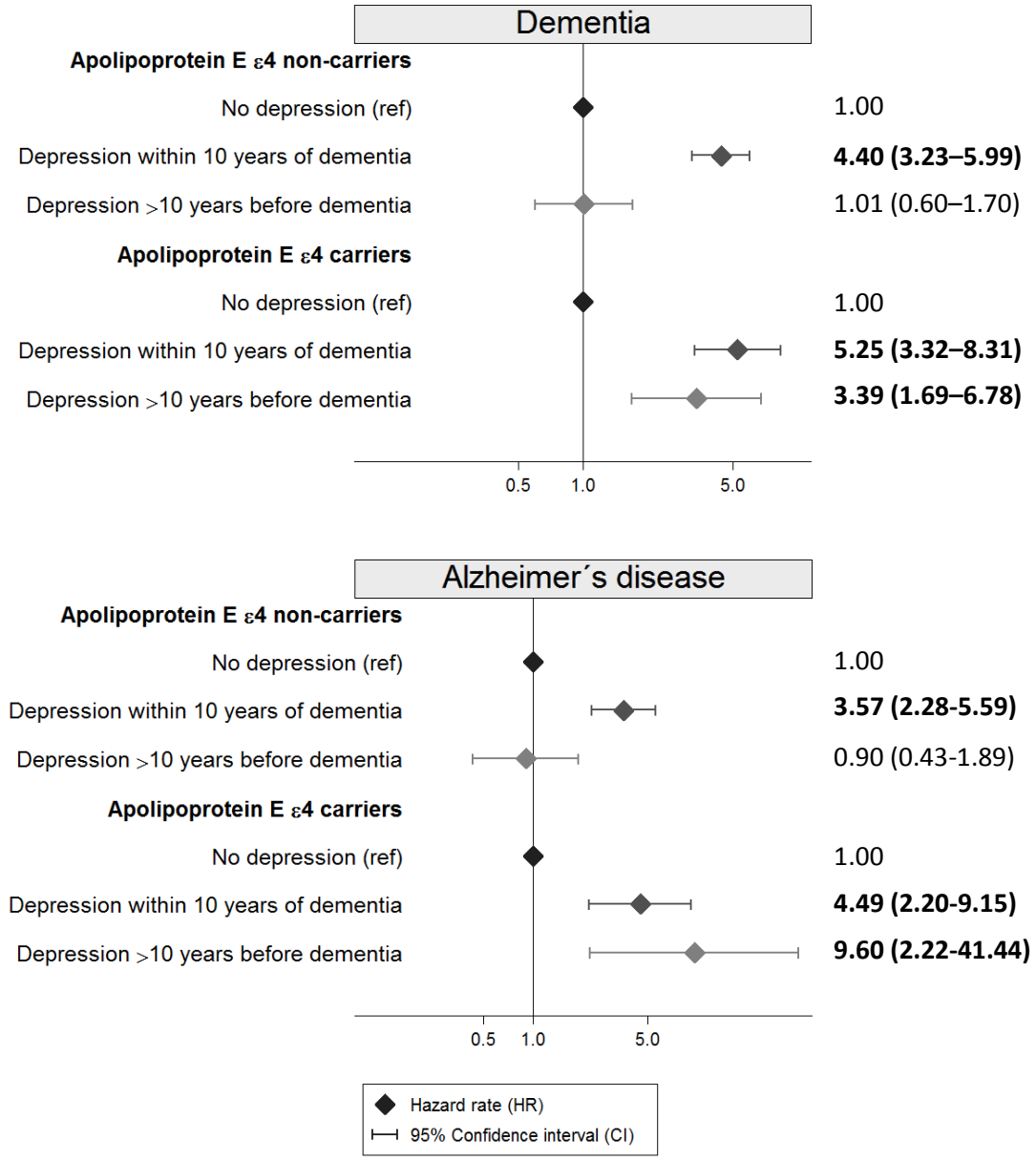


Figure 5: The association between depression and dementia, stratified by *APOE* genotype. Incidence rate ratios and 95% confidence intervals of dementia and Alzheimer's disease in the presence of depression in carriers and non-carriers of *APOE* ε4. Incidence rate of dementia in individuals with depression within ten years of dementia or more than ten years prior to dementia is compared to non-depressed individuals within the same genotype category. The models are adjusted for age, sex, sub-study, education, previous stroke, and sources of depression data available. *APOE*, apolipoprotein E.

Secondly, categorizing depression in relation to time to dementia onset led to a more marked difference in the association between the two depression categories and dementia than when depression was categorized in relation to age at onset. There was considerable overlap in the age range of depression within ten years and more than ten years prior to dementia. Hence, when using the age cut-off, some individuals with depression close in time to dementia onset are included in the midlife depression category, and some with depression more than ten years prior to dementia will be categorized as having late-life depression. The more marked difference for depression categorized in relation to time to dementia onset again indicates that the depression may very well be a prodromal feature occurring in the pre-clinical stage of dementia, and the association thus less dependent on age at onset.

7.2 Study II: No evidence of genetic overlap between CVD and dementia, but shared influences from lipids

The first part of this study focused on how genetic susceptibility to CAD influences the risk of dementia and its association with CVD.

The 13,231 individuals included in the study yielded 304,949 person years. During the study period, 1,430 individuals were diagnosed with dementia, of whom 868 had AD and 312 VaD. The mean age at dementia onset was 80 years. A total of 2,630 individuals were diagnosed with CVD at a mean age of 70 years.

The main findings from study II are summarized in Table 2. There was a 92% increase in the hazard rate of dementia during the first three years after a CVD diagnosis, which was reduced to normal after that. Similar findings have been shown in previous studies[77,89], but considering the long pre-clinical phase of dementia the nature of it is unclear. It is plausible that the CVD event acts as a stressor, leading to manifestation of dementia in susceptible individuals while those resilient are more likely to recover. The same effect was present in both AD and VaD, but stronger for VaD.

There was no association between the CAD GRS and dementia or its subtypes (hazard ratio (HR) 1.01, 95% CI 1.00–1.02 for dementia, HR 1.01, 95% CI 0.99–1.03 for AD, HR 1.01, 95% CI 0.98–1.04 for VaD). However, the GRS modified the association between CVD and dementia, such that the association was stronger in higher quartiles of the score (Table 2). This indicates that, although not having an increased risk of dementia overall, individuals with higher genetic susceptibility to CVD may be more susceptible to the stress induced by suffering from a CVD, and hence more likely to progress into dementia after such an event.

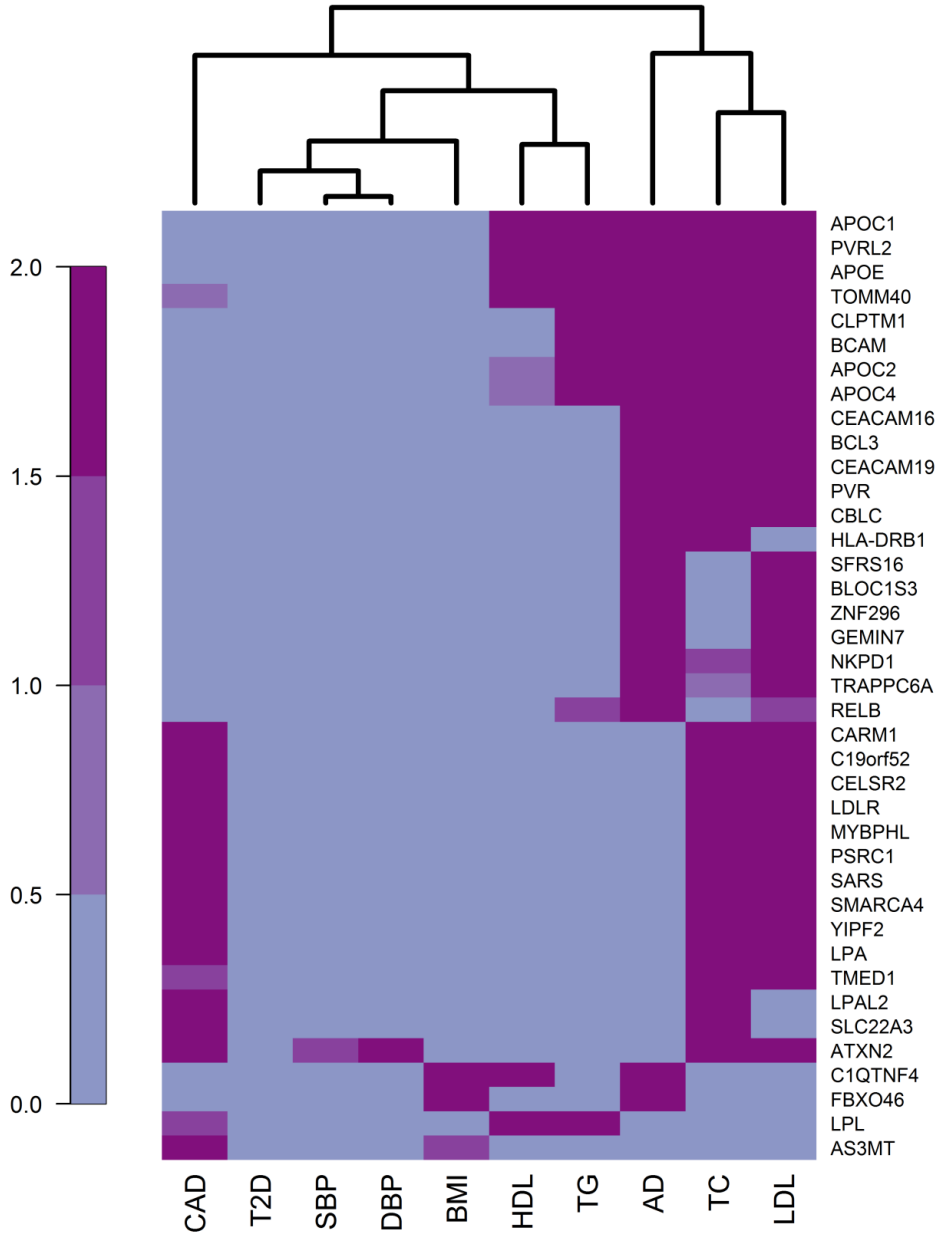
Table 2: Risk of dementia after a CVD diagnosis, for the total sample and stratified by quartiles of genetic risk score for CAD. Hazard ratios and 95% confidence intervals of dementia within the first three years after a CVD diagnosis and more than three years after a CVD diagnosis, for the total sample and stratified on genetic risk score for CAD. The model is adjusted for age, sex, education and type 2 diabetes during follow-up. CAD, coronary artery disease; CVD, cardiovascular disease.

	Total sample		Stratified on CAD genetic risk score								
			1st quartile		2nd quartile		3rd quartile		4th quartile		Trend p-value
All dementia											
First 3 y after CVD	**1.92**	**(1.57-2.36)**	**1.59**	**(1.05-2.41)**	**1.82**	**(1.19-2.78)**	**2.38**	**(1.65-3.42)**	**1.91**	**(1.28-2.86)**	**p<0.000001**
> 3 y after CVD	1.08	(0.92-1.26)	1.11	(0.81-1.51)	0.84	(0.61-1.18)	**1.34**	**(1.02-1.78)**	1.02	(0.74-1.40)	p=0.35
Alzheimer´s disease											
First 3 y after CVD	**1.47**	**(1.11-1.95)**	1.65	(0.97-2.79)	1.36	(0.75-2.48)	1.09	(0.56-2.12)	1.64	(1.00-2.70)	**p=0.02**
> 3 y after CVD	0.84	(0.67-1.05)	0.96	(0.63-1.49)	0.81	(0.52-1.25)	0.88	(0.58-1.35)	0.71	(0.47-1.08)	p=0.70
Vascular dementia											
First 3 y after CVD	**2.68**	**(1.85-3.89)**	1.30	(0.52-3.30)	1.80	(0.77-4.18)	**5.61**	**(3.24-9.71)**	**2.56**	**(1.08-6.08)**	**p<0.000001**
> 3 y after CVD	1.35	(0.99-1.83)	0.76	(0.37-1.57)	1.04	(0.57-1.89)	**1.99**	**(1.15-3.42)**	**1.92**	**(1.05-3.48)**	**p=0.01**

The second part of study II utilized GWAS summary statistics to investigate the shared genetic architecture between AD, CAD, and their common risk factors. As seen in Figure 6, no gene was significantly associated with both AD and CAD. However, both diseases had a significant number of genes in common with total cholesterol and LDL. Pathway analysis of the shared gene clusters identified 17 pathways for genes shared by AD and lipids, and 13 pathways for genes shared by CAD and lipids. Out of these, six of the pathways were identified for both AD and lipids and CAD and lipids, namely the statin pathway ($q = 5.82 \times 10^{-5}$ and $q = 9.88 \times 10^{-4}$), chylomicron-mediated lipid transport ($q = 7.95 \times 10^{-4}$ and $q = 4.36 \times 10^{-4}$), lipoprotein metabolism ($q = 1.40x10^{-3}$ and $q = 3.70 \times 10^{-5}$), retinoid metabolism and transport ($q = 2.48 \times 10^{-3}$ and $q = 1.50 \times 10^{-3}$), lipid digestion, mobilization, and transport ($q = 3.88 \times 10^{-3}$ and $q = 9.26 \times 10^{-5}$) and visual phototransduction ($q = 7.95 \times 10^{-3}$ and $q = 4.70 \times 10^{-3}$). This provides clues to the mechanisms by which the CAD GRS may modify the association between CVD and dementia. It is possible that individuals with a lipid dysregulation have higher risk of both CVD and dementia, as well as being more susceptible to develop dementia after a CVD. It should be mentioned that several of the SNPs included in the GRS for CAD reside in genes related to lipid levels[82], further indicating the importance of lipid regulation as a player in the association between CVD and dementia. Cholesterol is a well-established risk factor for CVD[90]. With the brain harboring 25% of the cholesterol in the body, and APOE functioning as a cholesterol transporter, it is also highly relevant to dementia[91]. However, since the brain has its own cholesterol metabolism, it is possible that different parts of the same pathways influence the risk of dementia and CVD.

A previous study that used LD-score regression to investigate genetic correlation across multiple phenotypes found no evidence of genetic overlap between AD and CAD[14]. In contrast to our findings however, BMI, T2D, blood pressure, and the lipid fractions all had a significant genetic correlation with CAD, while only HDL correlated with AD. One explanation for the discrepancies is that LD-score regression and the VEGAS method operate on different genetic levels. While LD-score regression utilizes information from all SNPs, VEGAS is focused on functional genes and will miss signals in the noncoding regions. In addition, VEGAS uses the p-values for each SNP while LD-score regression uses the β-values and the direction of it. In the latter, it may happen that small effects of opposite directions cancel out.

Taken together, the findings from study II indicate that there is an increased risk of dementia during the first years after a CVD diagnosis, and that this effect is stronger for genetically predisposed CVD compared to CVD with a lower genetic risk. However, the association is not due to genetic overlap, but may stem from shared influences from lipid dysregulation.



P-value for the genetic overlap with CAD	-	1	**0.05**	0.07	1	0.46	0.41	1	**$3x10^{-14}$**	**$1x10^{-13}$**
P-value for the genetic overlap with AD	1	1	1	1	**0.04**	**$1x10^{-4}$**	**$1x10^{-8}$**	-	**$1x10^{-15}$**	**$4x10^{-28}$**

Figure 6: Heat map and significance of genes associated with AD or CAD and their shared risk factors. Each row represents one gene, and each column one phenotype. Blue color indicates no/low significance while purple indicates high significance. The Fisher p-values for the number of overlapping genes are presented in the table. CAD, coronary artery disease; T2D, type 2 diabetes; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body-mass index; HDL, high-density lipoprotein; TG, triglycerides; AD, Alzheimer's disease; TC, total cholesterol; LDL, low-density lipoprotein.

7.3 Study III: Hypermethylation in the promoter region of *APOE* is associated with dementia

Among the 447 individuals with methylation information available, we identified 135 dementia cases, of whom 82 had AD. Eighty-four individuals were already diagnosed with dementia at time of blood draw, and 51 were incident cases. The mean age at diagnosis was 81 years. The sample included 181 complete twin pairs, of which 40 were discordant for dementia and 22 for AD.

We identified 205 individuals with a diagnosis of CVD, of which 112 were already diagnosed at time of blood draw and 93 were incident. Mean age at diagnosis was 78 years. Among the complete twin pairs, 69 were discordant for CVD.

Dementia and AD patients had significantly higher levels of methylation in the promoter region of *APOE* than controls (Figure 7). The signal was stronger in co-twin control analyses, showing that the effect is not due to genetic or shared environmental factors. No difference in methylation levels was detected in the other two regions of the gene. No other study has investigated the association between *APOE* methylation and dementia or AD in leukocytes, but some have done so using post-mortem brain samples with largely inconclusive findings. Two EWASs[38,39] and one study of methylation in AD related genes did not detect an association between *APOE* methylation and AD[86]. Conversely, one study of AD related genes[92] and one focusing on the CpG island in the 3´ exon of *APOE*[93] found a significant difference in methylation levels between cases and controls.

Interestingly, when analyzing prevalent and incident dementia separately we identified a stronger signal in incident (odds ratio (OR) 1.41, 95% CI 1.02-1.96) than prevalent (OR 1.25, 95% CI 0.97-1.62) disease. Some previous work has been done on circulating APOE levels. Although findings are inconsistent[91], two studies found an association between lower plasma levels of APOE and incident dementia[94,95]. Since promoter DNA methylation down regulates gene expression[96], these findings are in line with ours. There is evidence of correlation between APOE metabolism in the periphery and the CNS[94], and it is hence plausible that lower levels of circulating APOE either increase the risk of dementia or are an effect of the ongoing disease process. Although APOE levels both in blood and cerebrospinal fluid differ by *APOE* genotype[91], stratifying on genotype did not indicate a genotype specific effect for *APOE* methylation on dementia (interaction p-value 0.47 for region 1, 0.62 for region 2, and 0.99 for region 3). Results were similar for AD.

There was no association between *APOE* methylation and CVD, in the full sample or in co-twin control analyses (Figure 7). No differences were detected between prevalent and incident disease, nor was there any evidence of genotype specific effects (interaction p-value 0.40 for region 1, 0.97 for region 2, and 0.54 for region 3). To the best of our knowledge, only one previous study has investigated *APOE* methylation in CVD, finding no

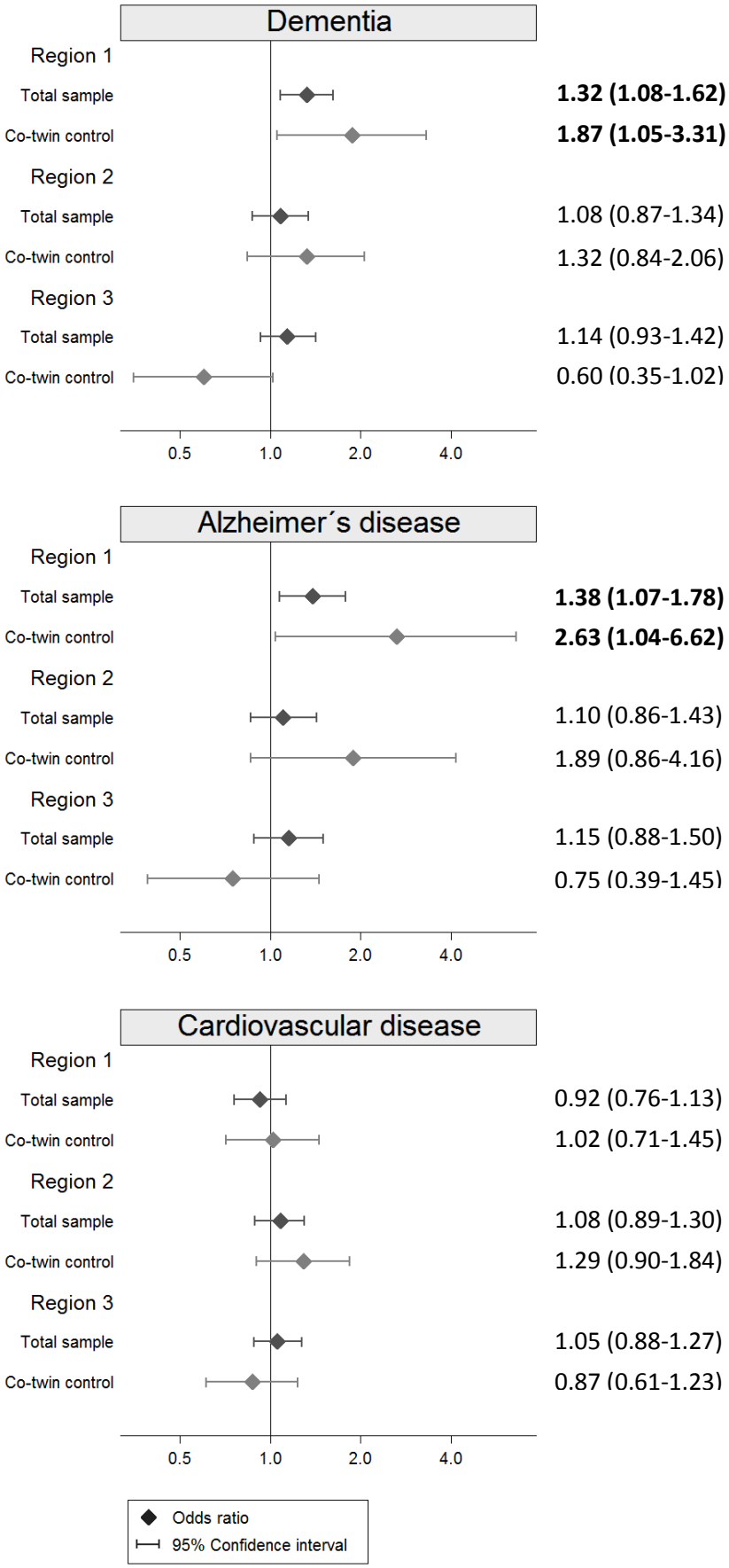


Figure 7: Odds ratio and 95% confidence intervals of dementia, AD, and CVD in relation to methylation in three regions of the *APOE* gene. Analyses of the total sample are adjusted for age, sex, smoking, and relatedness among twins, and co-twin control analyses for age and smoking.

difference in methylation of the promoter region in CAD patients compared to controls[97]. Studies of circulating APOE levels in relation to CVD are conflicting, but a review summarizing the findings for ischemic heart disease found indications that higher APOE levels are associated with a higher risk of disease[91].

7.4 Study IV: Methylation of *SORL1* and *BIN1* in leukocytes is associated with dementia and AD

Out of the 427 twins with methylation information from leukocytes, we identified 136 dementia cases, of which 85 were diagnosed with AD. The mean age at blood sample was 78 years, and mean age at dementia diagnosis 81. Among the 170 complete pairs included in the sample, 39 were discordant for dementia and 22 for AD.

Of the 39 twins with methylation information from neuronal cells, we identified 32 dementia cases, of which 21 had AD. Mean age at onset was 79 and mean age at death 88 years. According to the Braak staging, 18 of the twins had stage III or above, indicating presence of neurofibrillary tangles, while 17 had few or no neurofibrillary tangles.

The main findings from study IV are visualized in Figure 8. We detected a significant difference in leukocyte methylation of *SORL1* in dementia cases compared to controls (p=0.03) as well as in dementia discordant twin pairs (p=0.01). SORL1 is a receptor for APOE-rich lipoproteins, and there is evidence of decreased levels in the brains of AD patients[98]. One previous study has investigated methylation and expression levels of *SORL1* in relation to AD, but found no such association in neuronal cells or leukocytes [99]. Considering the strong association between *SORL1* and AD[98], this warrants further investigation.

There was a significant difference in leukocyte methylation of *BIN1* in AD discordant twin pairs (p=0.04), which did not reach significant at the α=0.05 level in the full sample (p=0.06). *BIN1* has been identified in both GWAS[8] and EWAS[38] of AD, and studies have shown increased expression levels of BIN1 both in the frontal cortex[100] and blood[101] of AD patients. Little is known about the mechanism through which the gene influences the risk of AD, but there is evidence it is mainly through interaction with Tau pathology[102].

We did not detect any difference in methylation level in any of the genes in cells from the PFC or CER, likely due to the substantially smaller sample size.

The CpG specific correlation between the three tissues varied across CpG sites, ranging between -0.62 to 0.94 with a median of 0.09. Similar variability has been shown in a previous study[103]. This, together with our findings in *SORL1* and *BIN1*, indicates that some methylation differences related to AD may indeed be present both in the affected neurons and in leukocytes.

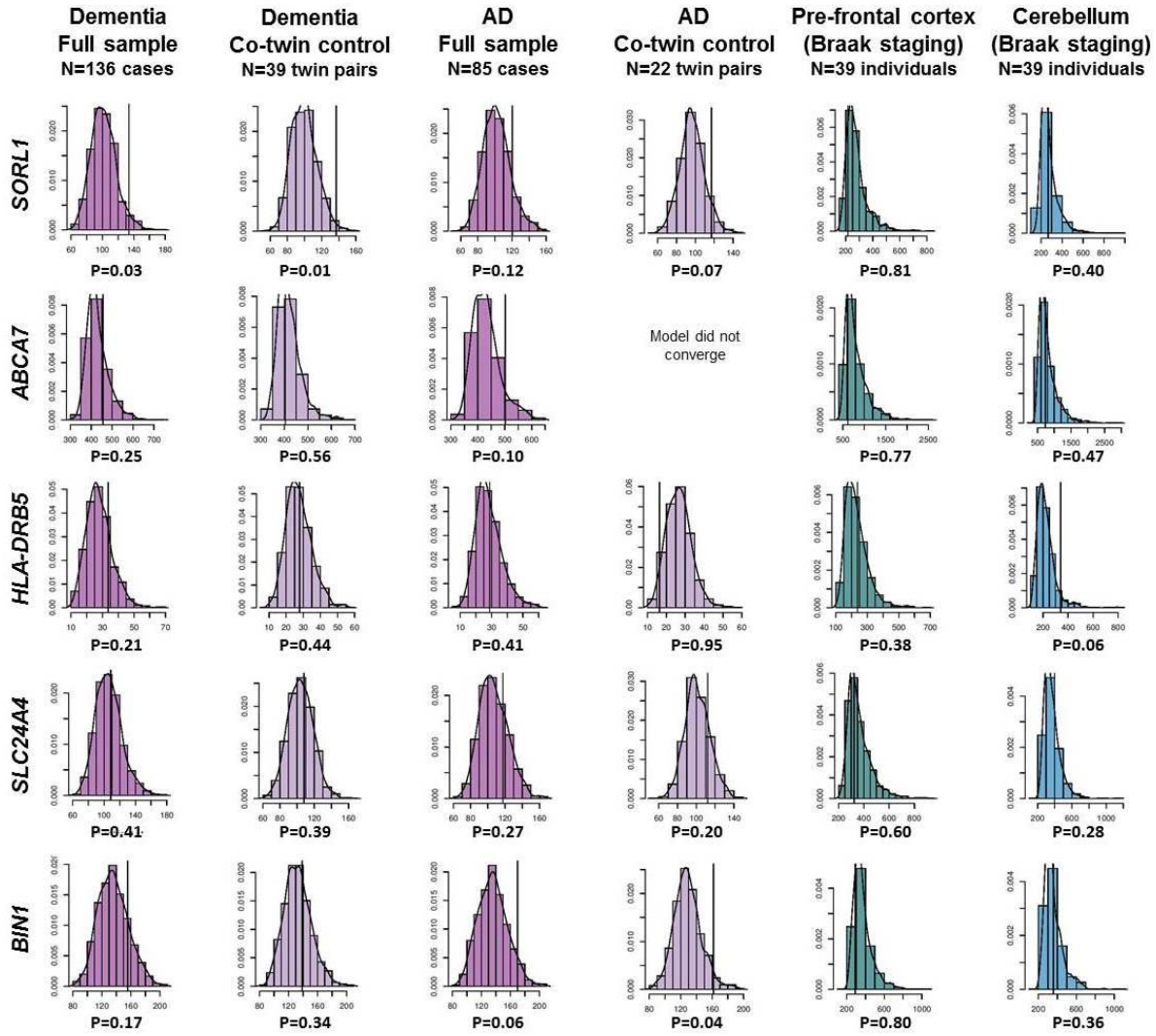


Figure 8: Gene-wide DNA Methylation in leukocytes in relation to dementia and Alzheimer´s disease, and in cells from the pre-frontal cortex and cerebellum in relation to Braak staging. Histogram and smoothed density function of the test statistics from randomly permutated data, with the vertical line representing the test statistic from the actual data. Logistic regression models of CpG M-values in relation to dementia and AD, and linear regression models of CpG M-values in relation to Braak staging. The full models were adjusted for age at blood sample, sex, and relatedness among twins. Models of neuronal cells were additionally adjusted for batch effects. The co-twin control models were adjusted for age at blood sample. P-values for each CpG site across the gene were combined into a test statistic using the Fisher product method, and the significance of the test statistic tested using random permutations. The co-twin control permutation test for *ABCA7* and AD did not converge. AD, Alzheimer´s disease.

7.5 Methodological considerations

7.5.1 Potential sources of systematic errors

The studies included in this thesis combine genetic information and robust statistical designs. They are based on well-established cohorts, with long follow-up time and rich phenotypic as well as genetic and epigenetic data. The possibility to follow participants through register linkage after end of follow-up further strengthens the studies. While twin status of the participants was not taken advantage of in the first two studies, it was of great importance for study III and IV. Nevertheless, as in all epidemiological studies, there are several methodological considerations that need to be taken into account.

Misclassification bias

The main limitation across these studies is identification of cases of dementia, depression, and CVD. Using data from registers enables us to retrieve information after the end of follow-up as well as information about diseases not assessed clinically as part of the studies, but there are several issues with this type of information. Since all diseases studied are primarily diagnosed outside the hospital setting and range from mild to more severe forms, it is possible that we only detect the more severe cases, as those more often require overnight stay at the hospital or specialist care. This is a substantial problem for depression diagnosis, where many cases would likely be missed in the registers. Indeed, very few cases were identified through this source in study I. The validity of dementia diagnoses from the NPR and CDR has been studied[104], and while the specificity is near perfect (>98%), the sensitivity is only 63% when both registers are combined. No such study has been performed on the PDR. Including information from all three registers likely increases the sensitivity, but there is of course a risk that the specificity is decreased due to prescription of dementia medications to patients with mild cognitive impairment. The low sensitivity would, as long as the misclassification is non-differential in relation to the exposure, bias the estimate toward the null. Differentiation between AD and VaD is likely a bigger issue and the subtype specific findings should be interpreted with caution.

In study II, we used a broad definition of non-stroke CVD to increase the power. However, we also ran sensitivity analyses using a strict definition of CAD, which only includes primary diagnoses of myocardial infarction or unstable angina. In the NPR, this definition has been shown to have validity above 95%[105,106]. Sensitivity analysis using this definition showed the same pattern as when the broader CVD definition was used.

There is of course a risk that having a register diagnosis for one disease increases the likelihood of receiving a diagnosis also of other diseases. This was mainly a concern in study II, where information about both exposure and outcome was extracted from the registers. If

this type of misclassification is present, it would inflate the estimates. However, using only clinically ascertained or the strict definition of CAD showed no indication of such a bias.

Another issue worth mentioning, not only with regard to using register data but also in general for diseases with a long preclinical stage, is estimation of the age at onset. It is likely that there is a delay between actual age at onset and register diagnosis. This further complicates associations as the diseases studied in this thesis begin years before manifestation. This of course needs to be considered when interpreting the findings from both study I and study II, since there is strong evidence of temporal effects.

Selection bias

Selection bias may arise when some individuals are more likely to participate in the study than others. This is certainly a problem in studies of aging, where individuals not only have to be healthy enough and willing to participate, but also to have survived long enough to be eligible. Furthermore, to be eligible for most of the sub-studies of aging used in this thesis, both members of the pair had to still be alive. This may lead to a study sample that is healthier than the general population and furthermore have a lower genetic susceptibility to severe diseases and mortality.

Residual confounding and reverse causation

We have done our best to adjust for any confounders, but the risk of residual confounding is always present in observational studies. Another issue is reverse causation, which is a major concern when studying late-life diseases with complex pathology and a long pre-clinical phase. Although certainly worth considering for study I and II, this is a chief concern in study III and IV. It is impossible to say whether the observed differences in methylation are a cause of dementia, or rather a consequence of the disease progress.

7.5.2 Ethical considerations

All participants included in the five sub-studies of aging provided informed consent, and the studies were approved by the Regional Ethics Board at Karolinska Institutet. Additionally, all data were pseudomized, using anonymous identifiers. As researchers, we have never had access to the personal identification number, name, or address of the participants. All data included in this thesis are stored on secure servers, where only those involved in the projects have access to the data.

In all research, the potential benefits must overweigh the cost for the participants, and another ethical issue is the discomfort the different testing phases may have induced. This is certainly not less of a problem in studies of elderly individuals, especially those with cognitive deficits. Collection of blood samples is of course directly associated with a certain amount of physical pain, but in addition, being interviewed and filling out questionnaires

may inflict some emotional distress, as many of the questions are of a personal nature. All participation is of course voluntary, and declination to participate must be respected. The time and effort made by the participants to contribute to this research is truly impressive, and we owe them much gratitude!

8 GENERAL DISCUSSION AND FUTURE DIRECTIONS

The studies included in this thesis focused on the role of genetic and epigenetic factors in dementia, and how they influence the associations between dementia, depression and CVD.

The findings from the first two studies highlight the complexity of late-life comorbidities, and the importance of taking both genetic factors and timing into account when studying these associations. The importance of considering timing when investigating risk factors for dementia has been shown not only for depression and CVD: e.g. higher BMI during midlife is associated with an increased risk of dementia[107], while high BMI in late life is associated with a lower risk[108]. The same pattern has been observed for high blood pressure[109]. It is plausible that all these factors during midlife do in fact increase the risk of dementia, but that they when measured during late life rather are a result of pre-clinical dementia progression. This also complicates investigations of genetic overlap and gene-environment interactions in late-life comorbidities. Even if genotypes are stable across the lifespan, their effect on dementia might differ with age.

Findings from the third and fourth study showed that not only allelic variation, but also methylation variation of genes related to AD is of importance for the disease. While methylation levels in the neuronal cells affected by disease are of great value to better understand the disease mechanism, the important limitation is of course that the tissue can only be collected post-mortem. Blood samples on the other hand, are easily accessible, and the possibility to collect samples over time opens up the possibility for both identification of biomarkers and a better understanding of disease progression. Findings from study IV indicate that methylation in some genes is in fact related to AD both in neuronal cells and in the PFC. This is indeed plausible since the effect of environmental factors may influence methylation throughout the body. Since dementia is complex and multifactorial, it may well be that methylation in blood is also relevant to disease risk and progression. In addition, there is a complex interplay between genetic variants and methylation, which would have the same influence on methylation across body systems.

Epigenetic factors may be the central in gene-environment interactions. Environmental factors throughout our lives have the potential to influence gene-specific as well as global methylation, and thereby affect the long term risk of disease[110]. However, the presence of meQTLs complicates interpretations as much of the effect of methylation may in fact be

driven by genotype. The co-twin control design is invaluable in this setting since it offers a natural way to adjust for the genetic factors. Interestingly, in both study III and study IV, the observed signals were stronger in the co-twin control analyses, despite a low number of discordant pairs. This further strengthens the finding, as it indicates that the association is not a result of genetic variants within the genes of interest.

Late-life diseases indeed have complex etiologies as well as associations. Many common risk factors such as smoking, low education, low physical activity, and unhealthy diet are associated with several late-life diseases, and it is likely that a similar pleiotropy can be found on the genetic level. The *APOE* gene is one such example. Other than the strong association with dementia, the gene is also associated with CVD, but in this case both the ε2 and the ε4 allele increases the risk of disease[111]. Studies of *APOE* and depression have been more inconsistent, but a meta-analysis concluded that, like for dementia, the ε2 allele decreases the risk of depression while the ε4 allele increases the risk[112]. However, the latter was true only for late-life depression, and it is hence plausible that the detected association is in fact with pre-clinical dementia. Not only is the *APOE* gene associated with all three diseases, there is also evidence of interactive effects. In addition to the interaction detected between *APOE* and midlife depression, the same pattern has been seen for CVD, with only ε4 carriers having a higher risk of dementia after a CVD diagnosis[77]. As methylation of the gene is also associated with dementia, the gene is an interesting target for further studies of gene-environment interactions mediated by methylation.

One reason for the lack of success in treatment development may well be that once symptoms appear the neurodegeneration has already advanced to a stage where it is too late to treat. Early detection of preclinical dementia would open up for new treatment strategies, where the disease process could be halted or slowed down before the disease reaches the advanced stage where symptoms appear. By delaying onset with as little as one year, the prevalence may be decreased by 9 million cases over the next 40 years[109]. This highlights the need for better identification of pre-clinical dementia. To achieve this, it is likely that a combination of biomarkers, genetic risk, epigenetic patterns, comorbidities, and prodromal symptoms will be most informative. Even if there is still a long way to go, a better understanding of these factors as well as of the interplay between them will help elucidate the disease mechanisms and hopefully help us avoid the expected increase in dementia prevalence.

9 CONCLUSIONS

I. Depression within ten years of dementia onset is associated with a higher risk of dementia regardless of *APOE* genotype, while depression more distal to dementia onset is a risk factor only in carriers of the ε4 risk allele.

II. Genetically predisposed CVD is more strongly associated with dementia than CVD with a lower genetic risk. The association between CVD and dementia is not due to shared genetic architecture, but may stem from shared influences from lipid dysregulation.

III. Hypermethylation of the promoter region of the *APOE* gene in leukocytes is associated with higher odds of dementia and AD, but not CVD. The effect is not dependent on *APOE* genotype, and remained when comparing dementia discordant twin pairs, which indicates it is not explained by genetic factors.

IV. In addition to carrying genetic variants associated with AD and being differentially methylated in cortical cells from AD patients, *SORL1* and *BIN1* are also differentially methylated in leukocytes collected pre-mortem from dementia and AD patients, respectively. Findings were stronger in discordant twin pairs, indicating the association is not due to genetic factors.

10 ACKNOWLEDGEMENTS

There is a long list of persons I would like to thank for help, support, stimulating discussions, and good company over the years:

First and foremost, I am truly grateful for having the best team of supervisors one could ask for!

Nancy Pedersen, my principal supervisor, for managing to always be there when I needed guidance and support while also giving me the freedom to work as independently as I wanted. You have an amazing ability to meet your students on our terms, and I could not have asked for a better supervisor.

Sara Hägg, my co-supervisor, you have always showed the utmost trust in me and my work. Despite having such a successful career, your care for your students always comes first. Other than being a fantastic supervisor, you are also an excellent role model – wise, honest, and open-minded.

Last but not least, **Alexander Ploner**, my statistical guru and superhero. As instrumental to the project as your statistical guidance has been, your support has been even more so. Thank you for always making me laugh, even at my darkest hours. Without you, I honestly could not have done this.

My former co-supervisors: **Anna Bennet Bark**, for first taking me to MEB. Even if we did not work much together after the first few months, I always felt that I had your support if needed. **Anna Dahl Aslan** and **Lotte Gerritsen**, you were much better supervisors to me than I was a student to you. I wish we had worked more together.

My other collaborators: **Chandra Reynolds**, your ability to work so hard and still be so warm and friendly is astonishing. You are a true inspiration! **Margy Gatz**, my constant and highly appreciated co-author. When you edit text, it is like magic happens! And thank you both for welcoming me to California and taking time out of your busy schedules to take care of me. I really had the best time! **Therese Andersson**, my first statistician, who explained how to interpret output so well that all the statistics courses were a piece of cake!

Patrik Magnusson, for being a breath of fresh air at the department, and for agreeing to chair my defense.

The awesome **Pedersen/Hägg** research group, which has come a long way since I first started here, in size, structure, collaborations, as well as socializing. **Adina Feldman, Andrea Foebel, Anna Dahl Aslan, Bojing Liu, Ci Song, Catalina Zavala, Dylan Williams, Fei Yang, Johanna Sieurin, Juulia Jylhävä, Kelli Lehto, Kathleen Bokenberger, Lotte Gerritsen, Miriam Mosing, Malin Ericsson, Xu Chen, Yiqiang Zhan, and Yunzhang Wang** - it has been a true pleasure working with you! And of course, thanks to **Nancy** and **Sara** for providing a stimulating work environment and excellent leadership. Past research administrators **Marie Krushammar, Erika Nordenhagen**, and **Michaela Rydén**, thank you for your patience and help with booking flight tickets, looking up grant numbers, and other more or less work-related issues. Thanks to **Adina, Ci, Dylan, Juulia, Kaavya, Kathleen, Kelli, Sara, Xu,** and **Yiqiang** for kindly inviting me to be part of your interesting projects, and to **Ci** and **Yunzhang** for helping me with mine. In addition, many of you deserve extra gratitude for voluntarily hanging out with me during and outside of office hours. In alphabetical order: **Catalina**, for being a constant ray of sunshine and the most amazing California guide. **Ci**, my second office mate, for being the only one as excited as me about molecular pathways. **Dylan**, for always being patient and polite. **Fei**, for thinking I have a clue about life (I honestly don´t know what gave you that impression) and, of course, for your hatred of walking. **Kelli**, for your witty comments and positive attitude. **Kaavya**, despite your absence your warmth lingers at MEB. **Malin**, my current office mate and moral compass. If just a proportion of the population were like you, the world would be a so much better place. **Yiqiang**, for trusting me and involving me in your projects. **Yunzhang**, for patiently helping me with everything related to methylation. And finally, a special thanks to **Kathleen**, my constant moral support. You seem so normal and nice, but underneath you are truly awesome! Thank you for keeping me sane through these years (bah).

My MEB friends outside of the research group. First and foremost, **Tong Gong**, my constant lunch-mate, third office mate, and dear friend. In retrospect, it is somewhat surprising that we became such good friends, considering how different our interests are. But I am so glad we did! **Emilio Ugalde Morales**, one of the most easy going and positive persons I know. Thank you for keeping me grounded! **Joanna Martin**, for being so strong-minded and smart, and at the same time so modest. Not to mention so much fun! Other friends and colleagues at the department, including but certainly not limited to **Andreas Jangmo, Anna Johansson, Bénédicte Delcoigne, Bronwyn Haasdyk Brew, Carolyn Cesta, Cecilia Lundholm, Chen Sou, Daniela Mariosa, Elisa Longinetti, Elisabeth Dahlqwist, Frida Lundberg, Gabriel Isheden, Hannah Bower, Iffat Rahman, Isabell Brikell, Jie Song, Johan Zetterqvist, Laura Ghirardi, Mark Taylor, Mina Rydell, Nelson Ndegwa Gichora, Qi Chen, Ralf Kuja-Halkola, Robert Karlsson, Sara Ekberg, Shuyang Yao, Vilhelmina Ullemar** – thank you for all the fun during various after-work events, and for all the fikas, lunches, and interesting discussions over the years. It really would not have been the same without you!

Thanks to the **BGA group** – always so much fun! And an extra thank you to the **Boulder bunch** for an amazing week of genetics and beer! The **lunch-bunch**, for making Thursdays the highlight of the week.

MEB is an excellent working environment, and I would like to express my gratitude to the **biostat, economy, HR, and IT department,** and of course the past and present **executive group,** for making all of our lives easier and for making MEB what it is. A special thanks to **Marie Jansson, Camilla Ahlqvist, Gunilla Nilsson Roos, Gunilla Sonnebring**, and **Katarina Ekberg**, for always helping out with a smile on your face. **Robert** and **Ralf**, thank you for patiently helping me with various aspects of genetic designs! To **PubMeb** for motivating us all to socialize after work. And, of course, to the **PhD group** for devoting your time to improve the working environment for us students.

My friends in the outside world – I consider myself lucky to have such a marvelous bunch of weirdoes as my friends. Thank you all for helping me remember to have fun. **Maria**, for being my everything! **Peter**, we have been through so much together, and I honestly can´t imagine what my life would be if we had not met. You will always be my ICE number one! **Erik**, thank you for painting me the beautiful cover illustration. **Susanne**, you have known me longer than anyone – 35 years and still going strong! Moreover, you turned out to be one of the wisest persons I know. **Malin** and **Pamela**, for always being ready to listen over a beer. **Andreas, Angelica, Anna, Helene, Lisa, Karro, Mats-Ola, Roberth, and Sandra**, for all the fun over the years!

To my **parents** and **siblings**, **brother**- and **sister in laws**, and many **nephews** and **nieces** – thank you for creating an environment both chaotic and stable. I am so grateful for having a family I actually enjoy spending time with. And an extra thank you to my **parents**, for always being proud of me, no matter what I do.

Margareta, my additional extra mother – for your extraordinary patience and humor.

My **parents in law,** for putting a roof over our heads when needed, and for always being there to help us.

And most important of all - my family. To **Marcus**, for somehow giving me both all the freedom and all the stability I could possibly ask for. And to **my children**, the true light of my life – without you, I´m nothing!

11 REFERENCES

1. Jylhava J, Pedersen NL, Hagg S. Biological Age Predictors. *EBioMedicine.* 2017.
2. Alzheimer's Association. 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association.* 2017;13(4):325-373.
3. Ferri CP, Prince M, Brayne C, et al. Global prevalence of dementia: a Delphi consensus study. *Lancet.* 2005;366(9503):2112-2117.
4. Wu YT, Beiser AS, Breteler MMB, et al. The changing prevalence and incidence of dementia over time - current evidence. *Nat. Rev. Neurol.* 2017;13(6):327-339.
5. Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. *Lancet.* 2006;368(9533):387-403.
6. Norton S, Matthews FE, Barnes DE, Yaffe K, Brayne C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *Lancet Neurol.* 2014;13(8):788-794.
7. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry.* 2006;63(2):168-174.
8. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* 2013;45(12):1452-1458.
9. Wen KX, Milic J, El-Khodor B, et al. The Role of DNA Methylation and Histone Modifications in Neurodegenerative Diseases: A Systematic Review. *PLoS One.* 2016;11(12):e0167201.
10. Rothman KJ. *Epidemiology : an introduction.* New York, N.Y.: Oxford University Press; 2002.
11. Plomin R, Plomin R. *Behavioral genetics.* 5th ed. New York: Worth Publishers; 2008.
12. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol.* 2017;18(1):77.
13. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry.* 2014;55(10):1068-1087.
14. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 2015;47(11):1236-1241.
15. Zaidi SK, Young DW, Montecino M, et al. Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. *Mol. Cell. Biol.* 2010;30(20):4758-4766.

16. Qazi TJ, Quan Z, Mir A, Qing H. Epigenetics in Alzheimer's Disease: Perspective of DNA Methylation. *Mol. Neurobiol.* 2017.

17. Purves D. *Neuroscience.* 3rd ed. Sunderland, Mass.: Sinauer Associates, Publishers; 2004.

18. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* 2016;17:61.

19. Gatz M, Fratiglioni L, Johansson B, et al. Complete ascertainment of dementia in the Swedish Twin Registry: the HARMONY study. *Neurobiol. Aging.* 2005;26(4):439-447.

20. Fotuhi M, Hachinski V, Whitehouse PJ. Changing perspectives regarding late-life dementia. *Nature Reviews. Neurology.* 2009;5(12):649-658.

21. Pimenova AA, Raj T, Goate AM. Untangling Genetic Risk for Alzheimer's Disease. *Biol. Psychiatry.* 2017.

22. Hung SY, Fu WM. Drug candidates in clinical trials for Alzheimer's disease. *J. Biomed. Sci.* 2017;24(1):47.

23. Ikram MA, Bersano A, Manso-Calderon R, et al. Genetics of vascular dementia - review from the ICVD working group. *BMC Med.* 2017;15(1):48.

24. Thal DR, Grinberg LT, Attems J. Vascular dementia: different forms of vessel disorders contribute to the development of dementia in the elderly brain. *Exp. Gerontol.* 2012;47(11):816-824.

25. Hugo J, Ganguli M. Dementia and cognitive impairment: epidemiology, diagnosis, and treatment. *Clin. Geriatr. Med.* 2014;30(3):421-442.

26. Smith EE. Clinical presentations and epidemiology of vascular dementia. *Clin. Sci. (Lond.).* 2017;131(11):1059-1068.

27. Meng X, D'Arcy C. Education and dementia in the context of the cognitive reserve hypothesis: a systematic review with meta-analyses and qualitative analyses. *PLoS One.* 2012;7(6):e38268.

28. Schubert CC, Boustani M, Callahan CM, et al. Comorbidity profile of dementia patients in primary care: are they sicker? *J. Am. Geriatr. Soc.* 2006;54(1):104-109.

29. Bennett S, Thomas AJ. Depression and dementia: cause, consequence or coincidence? *Maturitas.* 2014;79(2):184-190.

30. de Toledo Ferraz Alves TC, Ferreira LK, Busatto GF. Vascular diseases and old age mental disorders: an update of neuroimaging findings. *Curr Opin Psychiatry.* 2010;23(6):491-497.

31. Ridge PG, Hoyt KB, Boehme K, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol. Aging.* 2016;41:200 e213-220.

32. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science.* 1993;261(5123):921-923.

33. Poirier J, Davignon J, Bouthillier D, Kogan S, Bertrand P, Gauthier S. Apolipoprotein E polymorphism and Alzheimer's disease. *Lancet.* 1993;342(8873):697-699.

34. Bergem AL, Engedal K, Kringlen E. The role of heredity in late-onset Alzheimer disease and vascular dementia. A twin study. *Arch. Gen. Psychiatry.* 1997;54(3):264-270.

35. Lee C, Kim Y. Complex genetic susceptibility to vascular dementia and an evidence for its underlying genetic factors associated with memory and associative learning. *Gene.* 2013;516(1):152-157.

36. Schrijvers EM, Schurmann B, Koudstaal PJ, et al. Genome-wide association study of vascular dementia. *Stroke.* 2012;43(2):315-319.

37. Kim Y, Kong M, Lee C. Association of intronic sequence variant in the gene encoding spleen tyrosine kinase with susceptibility to vascular dementia. *World J. Biol. Psychiatry.* 2013;14(3):220-226.

38. De Jager PL, Srivastava G, Lunnon K, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* 2014;17(9):1156-1163.

39. Lunnon K, Smith R, Hannon E, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* 2014;17(9):1164-1170.

40. Mastroeni D, McKee A, Grover A, Rogers J, Coleman PD. Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease. *PLoS One.* 2009;4(8):e6617.

41. D'Addario C, Candia SB, Arosio B, et al. Transcriptional and epigenetic phenomena in peripheral blood cells of monozygotic twins discordant for alzheimer's disease, a case report. *J. Neurol. Sci.* 2017;372:211-216.

42. Magnusson PK, Almqvist C, Rahman I, et al. The Swedish Twin Registry: Establishment of a Biobank and Other Recent Developments. *Twin Res Hum Genet.* 2012;16(1):317-329.

43. Finkel D, Pedersen N. Processing Speed and Longitudinal Trajectories of Change for Cognitive Abilities: The Swedish Adoption/Twin Study of Aging. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 2004;11(2):325-345.

44. McClearn GE, Johansson B, Berg S, et al. Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science.* 1997;276(5318):1560-1563.

45. Johansson B, Whitfield K, Pedersen NL, Hofer SM, Ahern F, McClearn GE. Origins of individual differences in episodic memory in the oldest-old: a population-based study of identical and same-sex fraternal twins aged 80 and older. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 1999;54(3):P173-179.

46. Gold CH, Malmberg B, McClearn GE, Pedersen NL, Berg S. Gender and health: a study of older unlike-sex twins. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 2002;57(3):S168-176.

47. Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, Pedersen NL. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J. Intern. Med.* 2002;252(3):184-205.

48. Lichtenstein P, Sullivan PF, Cnattingius S, et al. The Swedish Twin Registry in the third millennium: an update. *Twin Res Hum Genet.* 2006;9(6):875-882.

49. Eriksson UK. *Inflammation-Associated Risk Factors for Alzheimer´s Disease and Dementia.* Karolinska Institutet; 2010.

50. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56-65.

51. Voight BF, Kang HM, Ding J, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* 2012;8(8):e1002793.

52. Radmanesh F, Devan WJ, Anderson CD, Rosand J, Falcone GJ, Alzheimer's Disease Neuroimaging I. Accuracy of imputation to infer unobserved APOE epsilon alleles in genome-wide genotyping data. *Eur. J. Hum. Genet.* 2014;22(10):1239-1242.

53. Davis S DP, Bilke S, Triche, Jr. T and Bootwalla M. methylumi: Handle Illumina methylation data. R package version 2.18.2. 2015.

54. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics.* 2013;14:293.

55. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;13:86.

56. Leek JT JW, Parker HS, Fertig EJ, Jaffe AE and Storey JD. sva: Surrogate Variable Analysis. *R package version 3.20.0.* *https://bioconductor.org/packages/release/bioc/html/sva.html*. 2016.

57. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* 2010;11:587.

58. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur. J. Epidemiol.* 2009;24(11):659-667.

59. Socialstyrelsen. *Kvalitet och innehåll i patientregistret: Utskrivningar från slutenvården 1964-2007 och besök i specialiserad öppenvård (exklusive primärvårdsbesök) 1997-2007.* 2009.

60. National Board of Health and Welfare. The National Patient Register. http://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish. Accessed July 2017.

61. Socialstyrelsen. *Dödsorsaksstatistik.* 2010.

62. National Board of Health and Welfare. The Causes of Death Register. http://www.socialstyrelsen.se/register/dodsorsaksregistret. Accessed July, 2017.

63. National Board of Health and Welfare. The Prescribed Drug Register. http://www.socialstyrelsen.se/register/halsodataregister/lakemedelsregistret. Accessed July, 2017.

64. Socialstyrelsen. *Fyra år med läkemedelsregistret* 2009.

65. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 1975;12(3):189-198.

66. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA.* 1993;269(18):2386-2391.

67. Gatz M, Reynolds CA, John R, Johansson B, Mortimer JA, Pedersen NL. Telephone screening to identify potential dementia cases in a population-based sample of older adults. *Int. Psychogeriatr.* 2002;14(3):273-289.

68. Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *Br. J. Psychiatry.* 1968;114(512):797-811.

69. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology.* 1989;39(9):1159-1165.

70. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders : DSM-III-R.* American Psychiatric Association; 1987.

71. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders : DSM-IV.* American Psychiatric Association; 1994.

72. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* 1984;34(7):939-944.

73. Roman GC, Tatemichi TK, Erkinjuntti T, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology.* 1993;43(2):250-260.

74. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl. Psychol. Meas.* 1977;1(3):385-401.

75. Himmelfarb S, Murrell SA. Reliability and validity of five mental health scales in older persons. *J. Gerontol.* 1983;38(3):333-339.

76. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am. J. Epidemiol.* 1982;116(3):547-553.

77. Eriksson UK, Bennet AM, Gatz M, Dickman PW, Pedersen NL. Nonstroke cardiovascular disease and risk of Alzheimer disease and dementia. *Alzheimer Dis. Assoc. Disord.* 2010;24(3):213-219.

78. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* 2015;47(10):1121-1130.

79. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518(7538):197-206.

80. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* 2012;44(9):981-990.

81. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011;478(7367):103-109.

82. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010;466(7307):707-713.

83. Liu JZ, McRae AF, Nyholt DR, et al. A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* 2010;87(1):139-145.

84. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* 2013;41(Database issue):D793-800.

85. Ma Y, Smith CE, Lai CQ, et al. Genetic variants modify the effect of age on APOE methylation in the Genetics of Lipid Lowering Drugs and Diet Network study. *Aging cell.* 2015;14(1):49-59.

86. Yu L, Chibnik LB, Srivastava GP, et al. Association of Brain DNA Methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 With Pathological Diagnosis of Alzheimer Disease. *JAMA neurology.* 2015;72(1):15-24.

87. Braak H, Alafuzoff I, Arzberger T, Kretzschmar H, Del Tredici K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* 2006;112(4):389-404.

88. Byers AL, Yaffe K. Depression and risk of developing dementia. *Nature Reviews. Neurology.* 2011;7(6):323-331.

89. van Oijen M, de Jong FJ, Witteman JC, Hofman A, Koudstaal PJ, Breteler MM. Atherosclerosis and risk for dementia. *Ann. Neurol.* 2007;61(5):403-410.

90. D'Agostino RB, Sr., Pencina MJ, Massaro JM, Coady S. Cardiovascular Disease Risk Assessment: Insights from Framingham. *Glob Heart.* 2013;8(1):11-23.

91. Rasmussen KL. Plasma levels of apolipoprotein E, APOE genotype and risk of dementia and ischemic heart disease: A review. *Atherosclerosis.* 2016;255:145-155.

92. Chibnik LB, Yu L, Eaton ML, et al. Alzheimer's loci: epigenetic associations and interaction with genetic factors. *Ann Clin Transl Neurol.* 2015;2(6):636-647.

93. Foraker J, Millard SP, Leong L, et al. The APOE Gene is Differentially Methylated in Alzheimer's Disease. *J. Alzheimers Dis.* 2015;48(3):745-755.

94. Wolters FJ, Koudstaal PJ, Hofman A, van Duijn CM, Ikram MA. Serum apolipoprotein E is associated with long-term risk of Alzheimer's disease: The Rotterdam Study. *Neurosci. Lett.* 2016;617:139-142.

95. Rasmussen KL, Tybjaerg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. Plasma levels of apolipoprotein E and risk of dementia in the general population. *Ann. Neurol.* 2015;77(2):301-311.

96. Pal S, Tyler JK. Epigenetics and aging. *Sci Adv.* 2016;2(7):e1600584.

97. Sharma P, Kumar J, Garg G, et al. Detection of altered global DNA methylation in coronary artery disease patients. *DNA Cell Biol.* 2008;27(7):357-365.

98. Yin RH, Yu JT, Tan L. The Role of SORL1 in Alzheimer's Disease. *Mol. Neurobiol.* 2015;51(3):909-918.

99. Furuya TK, da Silva PN, Payao SL, et al. SORL1 and SIRT1 mRNA expression and promoter methylation levels in aging and Alzheimer's Disease. *Neurochem. Int.* 2012;61(7):973-975.

100. Chapuis J, Hansmannel F, Gistelinck M, et al. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol. Psychiatry.* 2013;18(11):1225-1234.

101. Sun L, Tan MS, Hu N, Yu JT, Tan L. Exploring the value of plasma BIN1 as a potential biomarker for alzheimer's disease. *J. Alzheimers Dis.* 2013;37(2):291-295.

102. Tan MS, Yu JT, Tan L. Bridging integrator 1 (BIN1): form, function, and Alzheimer's disease. *Trends Mol. Med.* 2013;19(10):594-603.

103. Yu L, Chibnik LB, Yang J, et al. Methylation profiles in peripheral blood CD4+ lymphocytes versus brain: The relation to Alzheimer's disease pathology. *Alzheimers Dement.* 2016;12(9):942-951.

104. Jin YP, Gatz M, Johansson B, Pedersen NL. Sensitivity and specificity of dementia coding in two Swedish disease registries. *Neurology.* 2004;63(4):739-741.

105. Lindblad U, Rastam L, Ranstam J, Peterson M. Validity of register data on acute myocardial infarction and acute stroke: the Skaraborg Hypertension Project. *Scand. J. Soc. Med.* 1993;21(1):3-9.

106. Hammar N, Alfredsson L, Rosen M, Spetz CL, Kahan T, Ysberg AS. A national record linkage to study acute myocardial infarction incidence and case fatality in Sweden. *Int. J. Epidemiol.* 2001;30 Suppl 1:S30-34.

107. Dahl A, Hassing LB, Fransson E, et al. Being overweight in midlife is associated with lower cognitive ability and steeper cognitive decline in late life. *J. Gerontol. A Biol. Sci. Med. Sci.* 2010;65(1):57-62.

108. Dahl AK, Lopponen M, Isoaho R, Berg S, Kivela SL. Overweight and obesity in old age are not associated with greater dementia risk. *J. Am. Geriatr. Soc.* 2008;56(12):2261-2266.

109. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol.* 2011;10(9):819-828.

110. Tan Q, Heijmans BT, Hjelmborg JV, Soerensen M, Christensen K, Christiansen L. Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort. *Int. J. Epidemiol.* 2016;45(4):1146-1158.

111. Giau VV, Bagyinszky E, An SS, Kim SY. Role of apolipoprotein E in neurodegenerative diseases. *Neuropsychiatr. Dis. Treat.* 2015;11:1723-1737.

112. Feng F, Lu SS, Hu CY, et al. Association between apolipoprotein E gene polymorphism and depression. *J. Clin. Neurosci.* 2015;22(8):1232-1238.