BIOMARKER-AND PATHWAY-INFORMED POLYGENIC RISK SCORES FOR ALZHEIMER'S DISEASE AND RELATED DISORDERS

Danai Chasioti

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the School of Informatics and Computing,

Indiana University

May 2022

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Jingwen Yan, Ph.D., Chair

Andrew J. Saykin, Psy.D.

March 18, 2022

Kwangsik Nho, Ph.D.

Shannon L. Risacher, Ph.D.

Huanmei Wu, Ph.D.

DEDICATION

Dedicated to my parents that offer me the world.

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to my advisor Dr. Andrew J. Saykin, for providing me his invaluable expertise by allowing me to work by his side all these years. I will always remember our inspiring and constructive discussions, his patience, support, and guidance through this tough journey. I was blessed to work with one of the top scientists in the neuroscience field.

Of course, I cannot overlook the key role my co-advisor Dr. Jingwen Yan played in this work, which would not be possible without her directions and valuable feedback. Her assistance and advice were critical for the fulfillment of every research project I undertook. Special thanks to all the committee members for their valuable contribution and constructive criticism that helped me accomplish this thesis.

I would like to also thank all the colleagues and professors from the Indiana Alzheimer's Disease Research Center who supported me not only by sharing their domain expertise but also by providing me with invaluable data resources. I would like to personally thank Dr. Brenna C. McDonald, Dr. Kwangsik Nho, Kelly N.H. Nudelman and Dr. Shannon L. Risacher. Many thanks to Dr. J. Mandelblatt and all the collaborators from the Georgetown University. Last but not the least, I would like to thank Dr. Constantin T. Yiannoutsos and Dr. Li Shen for guiding me during the first steps of my carrier.

Finally, I would like to express my sincere gratitude to my dear family for believing in me, being always on my side, and supporting me at every step in my life. I could not have become the person I am, and achieve my highest goals without my mother Helen, my father George and my grandmother Eleftheria.

Danai Chasioti

BIOMARKER-AND PATHWAY-INFORMED POLYGENIC RISK SCORES FOR ALZHEIMER'S DISEASE AND RELATED DISORDERS

Determining an individual's genetic susceptibility in complex diseases like Alzheimer's disease (AD) is challenging as multiple variants each contribute a small portion of the overall risk. Polygenic Risk Scores (PRS) are a mathematical construct or composite that aggregates the small effects of multiple variants into a single score. Potential applications of PRS include risk stratification, biomarker discovery and increased prognostic accuracy. A systematic review demonstrated that methodological refinement of PRS is an active research area, mostly focused on large case-control genome-wide association studies (GWAS). In AD, where there is considerable phenotypic and genetic heterogeneity, we hypothesized that PRS based on endophenotypes, and pathway-relevant genetic information would be particularly informative. In the first study, data from the NIA Alzheimer's Disease Neuroimaging Initiative (ADNI) was used to develop endophenotype-based PRS based on amyloid (A), tau (T), neurodegeneration (N) and cerebrovascular (V) biomarkers, as well as an overall/combined endophenotype-PRS. Results indicated that combined phenotype-PRS predicted neurodegeneration biomarkers and overall AD risk. By contrast, amyloid and tau-PRSs were strongly linked to the corresponding biomarkers. Finally, extrinsic significance of the PRS approach was demonstrated by application of AD biological pathway-informed PRS to prediction of cognitive changes among older women with breast cancer (BC). Results from PRS analysis of the multicenter Thinking and Living with Cancer (TLC) study indicated that older BC patients with high AD genetic susceptibility within the immune-response and endocytosis

pathways have worse cognition following chemotherapy±hormonal therapy rather than hormonal-only therapy. In conclusion, PRSs based on biomarker- or pathway- specific genetic information may provide mechanistic insights beyond disease susceptibility, supporting development of precision medicine with potential application to AD and other age-associated cognitive disorders.

Jingwen Yan, Ph.D., Chair

Andrew J. Saykin, Psy.D.

Kwangsik Nho, Ph.D.

Shannon L. Risacher, Ph.D.

Huanmei Wu, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAO: Age at onset

ADNI: Alzheimer's Disease Neuroimaging Initiative

APE: Attention-processing speed-and-executive function

APOE: Apolipoprotein E

AUC: Area under the curve

BC: Breast cancer

CHD: coronary heart disease

CVD: Cardiovascular disease

CSF: Cerebrospinal Fluid

CN: Cognitively normal

Dem: Demented

EF: Executive function

EMCI: Early mild cognitive impairment

GWAS: Genome wide association study

HR: Hazard ratio

HRC: Haplotype Reference Consortium

HWE: Hardy-Weinberg equilibrium

ICV: Intracranial volume

LM: Learning and memory

LD: Linkage disequilibrium

LOAD: Late onset Alzheimer's disease

LMCI: Late mild cognitive impairment

MAF: Minor allele frequency

MEM: Memory

MS: multiple sclerosis

NGS: next generation sequencing

OR: Odds ratio

PCa: Prostate cancer

PD: Parkinson's disease

PHS: Polygenic hazard score

PRS: Polygenic risk score

RA: Rheumatoid arthritis

SMC: Significant memory concerns

SNP: Single nucleotide polymorphism

T2D: Type 2 diabetes

TLC: Thinking and Living with Cancer

WRAT4: Wide Range Achievement Test 4

Chapter 1

INTRODUCTION

1.1 POLYGENIC RISK SCORE IN PRECISION MEDICINE, ALZHEIMER'S, AND OTHER DEMENTIAS

The rapid advancement in genome sequencing technology, the innovative solutions for data storing, and the continuously improving computational power, have brought us closer to precision medicine than ever before, by enabling the collection and processing of enormous amounts of data. Precision medicine, aims to improve health by preventing, diagnosing, treating, or delaying the disease progress. Informed therapeutic decisions are made possible through individual risk assessment, based on environmental, lifestyle and genetic patient information. In contrast to lifestyle and age-related factors whose effect on a disease or a trait can be only assessed later in life, genetic information can be utilized at any point in life and support early disease prediction. While evaluation of a person's genetic predisposition is easier in the case of monogenic disorders, where single gene mutation is responsible for the disease development, that is not the case for complex disorders [1]. The majority of the human disorders are considered as complex, meaning that the disease risk is partly associated with multiple genetic variations, most of which individually account only for a small portion of an individual's overall genetic risk.

Alzheimer's disease (AD), one of the most common types of dementia, is an example of complex disease with high socioeconomic impact. It is a neurodegenerative disease, characterized by progressive deterioration in cognition, affecting the lives of patients

andtheir caregivers. According to recent data, in the U.S alone, more than 6 million people over the age of 65 are living with AD, the healthcare cost of which reached the $355 billion in 2021 [2]. With more deaths attributed to AD than breast cancer and prostate cancer combined, AD is considered as one of the leading causes of death [2]. As the pre-symptomatic AD diagnosis is currently very challenging, and no medications have yet proven to ameliorate the patients' cognition, it is important to focus on efforts that improve detection of high-risk individuals. Such efforts have the potential to advance public health by improving the quality of study cohorts and in turn the innovation in treatment development and biomarker discovery. On patient level, that could support decisions regarding dietary and lifestyle changes that could modify the lifetime-risk for AD by as much as 35%, according to recent studies [3]. Moreover, decision regarding treatment allocation and dosage/duration of medication can benefit by estimating the patient's genetic predisposition.

In complex diseases like non-familial AD, one of the challenges in determining a patient's genetic risk is identifying the (common) genetic variants that are contributing to the disease. Part of the perplexity of identifying new genes associated with the risk of sporadic AD, is that they are either rare mutations or they tend to have very small individual effect with exception those in *APOE* gene. Genome wide association studies (GWASs) have been traditionally used as one of the first steps in gene discovery. As solely experimental approach, GWAS's power to identify informative variants largely depends on the sample size. Fortunately, the current methodological and technological advancements have led to GWASs of substantially larger size compared to the past, which have greatly assisted the

variant discovery process [4]. Whereas association studies provide a great step in identifying candidate genetic variants, accounting for the overall genetic risk of these variants is a challenging task.

Polygenic Risk Score (PRS), which was by definition designed to account for the aggregated effect of multiple genetic variants, has gained much attention in the last decade [5]. The recent explosion in the availability of genetic and other biomedical data, has led to promising applications of PRS which in combination with other clinical measures could enhance the precision of disease prediction and diagnosis. Based on years of research, it is yet clear that "one PRS fits all" is not realistic and the decision on the best PRS highly depends on the data and the disease/trait of interest. Although the refinement of PRS is an ongoing research area with many potentials for improvement, the continuous release of new interesting biomedical data provides new application opportunities for PRS that could further support its efficiency and usefulness in promoting biomedical research.

1.2 OBJECTIVE AND AIMS

The overall goal of this work is the advancement of AD (and related dementias) research by introduction of biological relevant information in PRS. PRS is a mathematical formulation tallying the effects of multiple genetic markers through a single value, that expresses an individual's genetic liability for a disease. We embraced PRS as the main tool of this dissertation motivated by its unique characteristic to account for multiple genetic variants and thus, amplify the individually weak genetic effect on disease outcomes. This feature makes PRS an especially compelling tool for complex diseases like AD. In addition,

this decision was supported by the expanding evidence showcasing the potential of PRS to promote precision medicine when combined with other clinical measures.

This work is comprised by three aims. The objective of the first aim is to identify opportunities for novel applications and the potential for improvement in PRS. Through a rigorous literature review, based on 85 publications in several complex diseases especially focusing on AD, I present an overview that covers the PRS-related work, including methodological advancements, limitations, and applications of PRS. An extensive list of factors that can impact the performance of the score is provided along with a step-by-step guide for the PRS calculation. I provide an overview of findings that emanate by the PRS studies on several complex diseases and dedicate a section especially to AD.

The detailed appraisal of the PRS literature led to the realization that existing research is mainly focusing on PRSs derived by case-control studies, whereas there is a lack of approaches that integrate disease-specific biomarker. Taking this observation into consideration, the second aim focuses on the development of AD-related endophenotype PRSs and their assessment as predictive tool of AD outcomes of interest. Here, I utilize the *ADNI* study to answer the question of how individual AD-specific endophenotype PRSs perform compared to a combined PRS, in terms of various aspects of disease pathogenesis and progression. As a first step, I use selected AD biomarkers to retrieve four endophenotypes: amyloid (A), tau (T), neurodegeneration (N) and cardiovascular (V). Endophenotype GWASs are performed and used for the development of four individual endophenotype-PRSs (PRS_A, PRS_T, PRS_N, PRS_V) as well as one combined-PRS

(PRS_{ATNV}). A series of responses of interest, including dementia risk, dementia hazard, age at dementia diagnosis, and multiple biomarker trajectories, are studied for their association to each of the PRSs. Lastly, I provide a performance comparison between the individual and combined endophenotype PRSs. The findings point to the indication that, the level of genetic complexity and the implicated biological mechanisms of the responses that are linked to the combined PRS are different than these associated with the biologically more restricted individual endophenotype PRSs.

Driven by this observation, I hypothesize that if biologically targeted PRSs can enhance the prediction of specific biomarkers then, incorporating pathway-level information in a PRS, might help capturing changes in different cognitive domains. Specifically, for the third aim I utilize data from the *TLC* study to examine the potential link between AD-specific pathway-PRSs and the BC post-treatment changes in six cognitive domains over the span of three years. The rationale behind this study lies in the previously observed link between the two diseases that pinpoints to shared biological pathways. For the purpose of this project, I generate seven pathway-PRSs based on genome-wide significant SNPs that have been previously strongly linked to AD and have been enriched to seven distinct biological pathways. Further association analysis examines the potential role of the aforementioned scores in the treatment-related cognitive changes of six domains, both cross-sectionally and longitudinally. The results returned significant association of post-treatment cognitive performance of older BC survivors with genetic risk linked to immune and endocytosis pathways. Each of these pathways was associated with a different set of cognitive domains. They also suggested worse executive function and visuospatial ability

for participants that were following chemotherapy±hormonal therapy rather than hormonal-only therapy.

1.3 SIGNIFICANCE

Determining an individual's genetic susceptibility in complex diseases is not a trivial task as multiple variants are involved, each contributing a small portion of the overall risk. In addition, Alzheimer's disease is characterized by phenotypic heterogeneity and highly unstable progression which further complicates the efforts early prediction and treatment development. Traditionally, case-control PRSs have been developed and utilized as tools of risk assessment, but it is challenging for such scores to provide any further insights regarding the disease's pathogenesis and progression.

This is the first study to investigate the potential of endophenotype-PRSs and its ability to ameliorate our understanding of complex diseases such as AD. As far as I know, currently there is no other publication about developing or studying AD-related endophenotype-PRS. Here it was shown that individual endophenotype-PRSs are beneficial for responses like amyloid and tau that are linked to specific biological function, but for responses that involve multiple biological pathways, like dementia risk and neurodegeneration, a combined/overall PRS is preferred. That information is important for several reasons. First, it suggests that there is no universal PRS that performs equally well in predicting every outcome of interest. Second, indicates that the biological function of the PRS SNPs is important and needs to be accounted for depending on the outcome one wish to study. This contradicts the established practice of using a single risk-based polygenic score (either based on the odds or hazard of the disease) for prediction of any type of disease outcome.

Based on data from older BC patients, it was also found that different cognitive domains are associated with dementia-related genotypic information. In the past, it has been observed that *APOE* risk can contribute to post-treatment cognitive impairment among BC survivors. In this work it was observed that, changes in executive function and visuospatial abilities are linked to immune-response genetic risk, whereas memory and language abilities are related to endocytosis-PRS. This further confirms that different type of information can be captured by aggregated genetic scores that embed SNPs with distinct biological characteristics. It also provides additional information regarding the biological function of the dementia-related genes that are possibly implicated in cognitive changes of BC individuals. Finally, it suggests mechanistic pathways that can be used to enhance the understanding of the cancer-induced cognitive problems.

1.4 CONTRIBUTION

The present work recognized the potentials of utilizing biomarker information in the development of polygenic risk scores, for promoting AD research. It offers opportunities for generating and studying novel hypothesis on the role of endophenotype-PRS in AD and other complex diseases. The synopsis of the existing PRS methods and the presentation of the elements that affect PRS's predictive accuracy, can be utilized to support future efforts for methodological refinement and implementation. Furthermore, it can be of use to investigators that are interested in utilizing PRS to support their research. Endophenotype-specific SNP weights derived by endophenotype GWASs, reflecting a risk which is biomarker-related rather than disease-status-related. Although AD biomarker GWASs

exist, endophenotype GWASs that represent a broader class of biomarkers, have not been previously reported. In addition to biomarkers that fall in the widely adopted A/T/N classification scheme, a cardiovascular biomarker was also considered in the PRS development. To the best of my knowledge, no other work has ever developed or studied the vascular-PRS before. The fact that both the endophenotype- and pathway- PRSs that were studied here, consist of SNPs with common biological characteristics, makes them highly interpretable. That could be a much-desired characteristic for PRS depending on the research question. The results obtained from pathway-PRS on BC survivors, can be utilized toward understanding the cancer-induced changes in different cognitive domains and assist with treatment decisions. Finally, another important remark that resulted by this dissertation the need for generating larger datasets that contain an extensive collection of AD biomarkers comparable to *ADNI*.

Chapter 2

PROGRESS IN POLYGENIC COMPOSITE SCORES IN ALZHEIMER'S AND OTHER COMPLEX DISEASES

In this review chapter we will explore how polygenic approaches that incorporate the aggregate influence of multiple genetic variants can contribute to a better understanding of the genetic architecture of many complex diseases and facilitate patient stratification. Polygenic risk scores (PRS) can serve as tools which combined with other clinical measures, could enhance clinical study designs through enrichment strategies. This review addresses polygenic concepts, methodological developments, hypotheses, and key issues in study design. PRSs have been applied to many complex diseases and here we focus on Alzheimer's disease (AD) as a primary exemplar.

2.1 POLYGENIC LANDSCAPE OF COMPLEX DISEASES

The hypothesis of multifactorial etiology of complex diseases has its roots in Fisher's 1918 quantitative demonstration that human variability in traits such as height and other biometric characteristics can be explained by the additive effect of multiple genetic factors [6]. In contrast to the single-gene etiology of Mendelian diseases, complex diseases are influenced by multiple gene variants and environmental factors [7]. The individual effects of these variants are usually very small [8] making determination of the genetic architecture of complex diseases challenging. Combinatorial genetic metrics such as the polygenic risk score (PRS) and its variations are designed to address these challenges.

The PRS expresses the cumulative genetic risk for an individual as an additive function of

the effect of each genetic marker. Polygenic methods have been widely utilized to investigate many diseases, e.g., congenital malformations [9], breast cancer (BC) [10, 11], type 2 diabetes (T2D) [12], schizophrenia and other psychiatric disorders [13, 14], and Alzheimer's disease (AD) [15, 16]. Use of PRS for risk stratification and classification is contributing toward the goals of precision medicine. This is enabled by advances in high-throughput genotyping and next generation sequencing (NGS) and the availability of large-scale genome-wide association studies (GWAS), which continuously expand the list of disease-related genetic markers [17]. Additional PRS applications include patient stratification [12, 15, 18, 19], exploration of genetic architecture [11, 20, 21], and studies of genetic overlap between traits [10, 13, 22].

Several review articles have been dedicated to facets of research on PRS [22-26]. Wray et al. [26] discussed some of the methodological aspects that influence PRS in the context of psychiatric disorders. Mistry et al. [24] systematically reviewed the association of schizophrenia-related PRS with different phenotypes; others mainly focus on disease-specific findings (e.g., [22, 23, 25]) or do not examine methodological factors related to the development and application of PRS.

Here, we review key methodological issues to assist researchers interested in employing PRS for studies of complex disease and for clinicians interested in potential future clinical applications in precision medicine. We overview the state-of-the-art methods for PRS construction and discuss study design and disease characteristics related to performance. Finally, we provide an overview of the contributions of PRS to a wide spectrum of diseases and a detailed overview of applications to Alzheimer's disease.

2.2 CALCULATION OF POLYGENIC COMPOSITE SCORES

By combining the influence of each single nucleotide polymorphism into a single measure, the PRS represents the aggregate influence of the genetic variation. There are two approaches for PRS calculation: 1) simple sum of SNPs, and 2) weighted sum of SNPs (Figure 2.1). The first approach [10, 12, 27, 28] assumes an equal contribution of all SNPs to disease risk and is rarely realistic as some variants carry a much larger contribution to disease heritability (e.g., the *APOE* ε4 allele in AD [29]). In the weighted sum approach, each SNP is weighted by its estimated disease effect size, therefore accounting for its unique contribution to disease risk or outcome [10-16, 18-20, 30-57]. Next, we discuss SNP selection and weight estimation in more detail as these are two important methodological aspects of developing weighted PRS.

2.3 SNP SELECTION

The selection of candidate SNPs is critical because these variants constitute the building blocks of the PRS. A simple strategy is to retain all the SNPs without filtering. This may be effective for genetically underexplored diseases or diseases with many small to moderate SNP effects. However, the PRS's performance may suffer by incorporating many non-informative or very weakly associated SNPs. Alternatively, one can retain a subset of SNPs based on predefined criteria (e.g., those passing an arbitrary p-value threshold in the GWAS results).

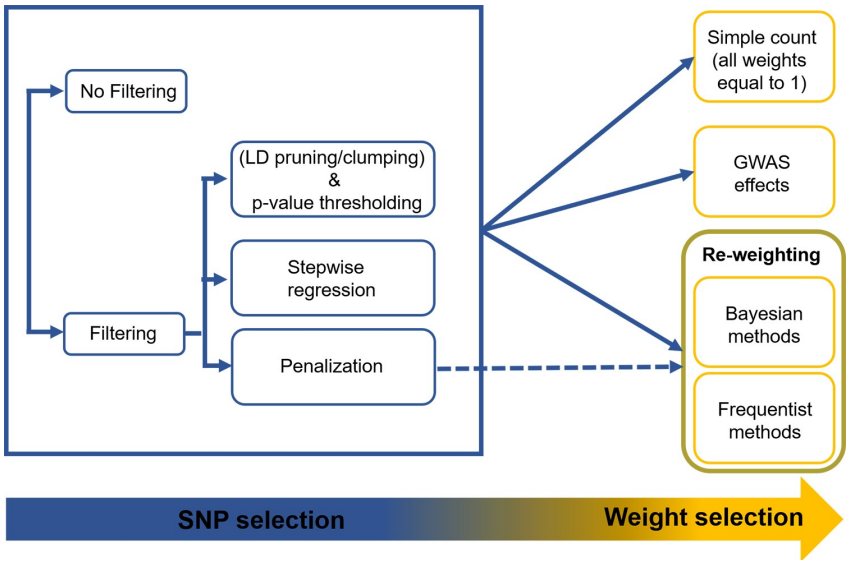

Figure 2.1: Polygenic risk score calculation. Step1) SNP selection, Step2) weight calculation: After selecting candidate SNPs for the score (with or without filtering), one can choose to not assign any weights to the SNPs (PRS is a simple sum of SNP alleles), to use the existing GWAS effect sizes as SNP weights, or to re-calculate the weights (re- weighting). In the case of re-weighting, new weights can be estimated by including the SNPs in a regression model (e.g., Cox). The resulting effect estimates will be the new weights for the PRS calculation. Penalization techniques (either frequentist e.g., Lasso or Bayesian e.g., LDpred) can also be used for re-weighting. These methods can achieve SNP selection and weight estimation simultaneously, by setting some of the SNP weights to zero. Penalization methods can be either applied on the filtered or on the original SNP list.

This ad-hoc cut-off selection, however, may omit some informative markers with small effect size. Thus, the PRS-disease association may significantly vary under different thresholds [13, 35, 53].

Another challenge is redundancy of informativeness of variants, especially in the case of linkage disequilibrium (LD) where nearby SNPs have highly similar associations without adding further explanatory power. This can be addressed by SNP filtering techniques such as LD pruning followed by p-value thresholding. The majority of the SNPs in a LD block are removed by random pruning or clumping. The remaining SNPs are further filtered by thresholding their p-values. PRSice is an example of a software approach employing LD pruning for automated calculation of the PRS [58]. It allows SNP selection under a range

of p-value thresholds offering a more precise cut-off choice. One caution is that overfitting issues may arise based on threshold selection criteria [59, 60].

Stepwise regression can also be used for SNP selection [15, 49, 55, 56]. In this approach, a SNP is retained based on whether it significantly improves the model's predictive ability. This purely statistical approach has the disadvantage of ignoring prior knowledge of LD structure and possible disease-variant relations.

2.4 SNP-WEIGHT CALCULATION

Another key factor for PRS performance is the choice of SNP weights. GWAS-derived statistics or risk estimations on an independent sample are commonly used as PRS weights [11, 15, 49, 56, 61]. A polygenic hazard score (PHS) extension of this approach that has been promising in AD research was reported by Desikan et al [11, 15, 49, 56, 61]. The PHS is also derived as a weighted sum of SNPs but in this case each SNP's weight is expressed by a hazard ratio (HR) estimated using a survival model where SNPs are entered as predictors.

GWAS genotypes in a PRS discovery sample may not be sufficiently representative of those in the validation or application set leading to attenuated performance of the PRS. Other factors that influence performance are LD and regression to the mean or "winner's curse". Adjusting SNP weights may help address these concerns. Next, we consider the two main approaches to optimized SNP re-weighting: 1) those based on Bayesian inference and 2) those based on frequentist inference.

LDpred [62] uses known LD structure as a prior to derive new SNP weights, without requiring raw genotype data or p-value thresholds. When applied on simulation data, LDpred demonstrated improved trait prediction accuracy compared to traditional methods without LD information [62]. AnnoPred [63] further improved LDpred, by assuming that each SNP's biological identity contributes to the SNP-specific heritability. With this additional assumption and tested on 5 diseases, AnnoPred achieves higher precision in weight estimation (using functional annotation as a prior), better prediction accuracy of disease status, and better risk stratification ability, compared to LDpred [63]. Another Bayesian based method [44] is the *doubly-weighted* PRS, which addresses the "winner's curse". It weights each SNP by both its estimated effect on the trait and the probability that its p-value is less than a cut-off. In a study of prevalent T2D, inclusion of the *doubly-weighted* PRS in a logistic model showed significantly better fit than the model with the conventional GWAS-based weighted PRS. Although evidence was not presented in their study, the authors propose that, their method reduces "winner's curse" bias compared to the conventional GWAS-based weighted PRS. The efficiency of the aforementioned methods is highly dependent on parameter tuning. An alternative Bayesian method that requires no parameter tuning [60] corrects a SNP effect by utilizing GWAS *z-statistics* and by assigning a probability for the SNP being not causal.

Frequentist approaches, including shrinkage regression (e.g., Least Absolute Shrinkage and Selection Operation (Lasso) [64]) and linear mixed models (LMM) (e.g., GeRSI [65]), have also been utilized for PRS calculation. Shrinkage methods, which penalize the SNP effect estimates to avoid overfitting, show higher precision and power, compared to

univariate tests [66]. They can successfully handle LD, SNP interactions, and non-genetic covariates [67]. Lasso estimates minimize the sum of squared residuals and assign a penalty on the absolute sum of the predictors' coefficients. Hence, less informative predictors are assigned smaller weights or removed from the model. Lassosum [68] is an example that applies a Lasso-type formula for SNP effect estimation. Despite the need for parameter tuning, it is computationally appealing and outperforms both pruning-thresholding and LDpred methods [68]. LMM, by contrast, treat the most significant SNPs as fixed effects with regard to disease status, and less significant SNPs as having random effects [65]. Here, the fixed effect SNPs are treated as parameters that need to be individually estimated, whereas the random effect SNPs do not require individual estimation since they are considered to be random variables with a common distribution. Both methods, however, are based on distributional assumptions of the genetic effects. Specifically, the shrinkage methods assume a skewed effect distribution, where the majority of the SNPs have small effects and only few have large effects; LMM assumes a normal distribution of effects. If these assumptions are violated, the PRS performance may suffer. To overcome this issue, "hybrid" methods such as Bayesian sparse linear mixed model'' (BSLMM) [69, 70] and LMM-Lasso [69, 70] were developed that combine the LMM and regularization methodologies.

Both Bayesian and frequentist methods can be further improved by embedding non-genetic information. For example, Sleegers et al [71] used age-specific odds ratio (OR) as *APOE* weight in the PRS and showed significantly improved discriminative power compared to a simple weighted score. The above referenced Desikan et al [15] PHS approach employed

age-specific PRS weights. Although SNP selection and weighting are key elements in PRS performance, other factors also play an important role. We next consider factors influencing power and accuracy.

2.5 POWER AND ACCURACY OF POLYGENIC COMPOSITE SCORE

In addition to the methodological factors discussed above, PRS performance is also influenced by study design. Here, we describe some of the factors that could negatively influence analysis results (Fig 2.2). Power and accuracy are positively correlated with sample size [63, 72]. For a highly heritable and loosely defined trait (e.g., heterogeneous psychiatric disorders), the sample size required to achieve the maximum possible area under the curve (AUC), is significantly smaller than that for a less heritable but strictly defined trait (e.g. specific cardiovascular diseases such as myocardial infarction or stroke) [17]. For less strictly defined diseases with larger the sample sizes, more "liberal" SNP p-value inclusion thresholds are required [73]. However, heterogeneity problems may arise as the sample size increases [17]. An alternative strategy to improve power is by varying the p-value thresholds for SNP selection. The optimal p-value threshold is determined by the underlying genetic architecture of disease and the sample size [17, 73]. For example, loosely defined traits will benefit more by a relaxed p-value threshold, compared to strictly defined traits (e.g., diseases with a small number of informative SNPs) [72] because the heritability of loosely defined traits spreads among a larger number of genetic markers and a relaxed cut-off allows more heritability to be explained. However, threshold increases should be made cautiously as these are usually accompanied by increases in Type I error and reduced power [73], and may lead to biased effect estimates with high levels of LD. In

contrast, a strict p-value cut-off will be more beneficial for strictly defined traits by eliminating non-informative SNPs [17].

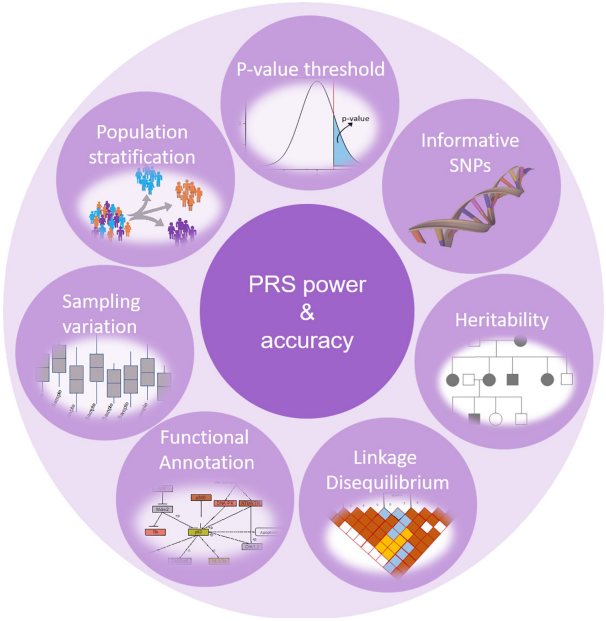

Figure 2.2: Factors affecting PRS accuracy. Disease related factors (e.g., heritability, functional annotation, LD structure, and number of informative SNPs) as well as study design aspects (e.g., sample size, p-value threshold for SNP selection, and sampling variability), can affect the power and performance of the PRS. Given the hypothesis tested and the disease characteristics, the best PRS performance can be achieved by identifying the appropriate sample size, SNP selection threshold and LD handling approach.

In some cases, the desired AUC for a given trait cannot be achieved using only genetic data and incorporation of additional information (e.g., functional annotation of the PRS markers [63] and pathway specific PRS [74]) may be beneficial.

As in all research, study goals should be clearly and operationally defined. Since PRS is used either for association analysis or for individual prediction, the sample requirements vary in each case. Dudbridge [17] discusses sample size and power and suggests that sample sizes are adequate to ensure a well powered association study when independent datasets for training and testing are available. If the latter is not possible, 1:1 splitting ratio

between the two sets is advised [17]. In contrast, individual prediction requires a significantly larger training set compared to the testing set [17]. PRS may be unable to successfully discriminate risk groups when there are limited training sample sizes which attenuates precision in the PRS-explained variation [17].

Additionally, false positive results can occur from the presence of population stratification, due to systematic genetic differences among populations [13]. In multi-site studies or multi-ethnicity samples, the population structure should be controlled to avoid such bias.

2.6 POLYGENIC RISK SCORE APPLICATIONS

Existing research using PRS mainly focuses on two problems: 1) association analysis and 2) outcome prediction. Although use of PRS has not achieved clinical accuracy levels yet, use of such composites has led to some interesting discoveries and shown potential in diseases like cancer [10, 11, 33, 47, 55], psoriasis [18], rheumatoid arthritis (RA) [18], multiple sclerosis (MS) [30], mental disorders [13, 14], atherosclerosis [46], T2D [12, 28, 33, 44], asthma [27], Parkinson's disease (PD) [20, 41], and cardiovascular diseases (CVD) [19] including coronary heart disease (CHD) [32].

Association analysis quantifies the relation between two sets of features such as phenotype and genotype (e.g., SNPs). In this context, PRS is used to assess the differential biology between disease types or stages [11, 14, 48], to identify risk strata [18], to assess treatment response [46] and to identify genetic overlap between diseases [10, 13]. Association of a simple sum PRS with T2D genetic risk strata found that, men and women in the highest PRS quantile had ~2.8 and ~2.2 times higher risk of developing T2D respectively,

compared to those in the lowest PRS quantile [12]. Similar findings were reported with a GWAS weighted PRS [12]. Another study showed that adopting a healthy lifestyle can reduce the CVD risk, regardless of the individual's genetic background [19]. For participants with high genetic risk, those with healthy lifestyle had 46% lower risk of CVD compared to those with unhealthy lifestyle [19].

The PRS has also been employed to explore genetic overlap between different diseases, where the PRS derived from one disease is evaluated on another disease. Motivated by this, a *multi-polygenic score* [75] was proposed recently, where multiple PRSs from different GWASs are combined for outcome prediction. Compared to a single PRS, this method explained more variability when applied to three traits (i.e., BMI, educational achievement and cognitive ability). Its use is advised for situations with modest sample size [75].

As a tool for individual prediction, PRS has also shown potential in screening process. For example, in a study on aggressive prostate cancer (PCa) and using polygenic hazard score (PHS) it was observed that, males in high genetic risk (>98th centile) have almost triple PCa hazard, compared to those in average genetic risk [55]. For PCa patients who had undergone radical prostatectomy, PCa recurrence was predicted with AUC= 88.8% [47]. Moreover, the 10-year recurrence-free rate for those in high genetic risk is almost half (46.3%), compared to people in the lowest genetic risk group (81.8%).

The use of PRS in public health and medicine has significant potential. The above results indicate the potential role of PRS to serve as a biomarker relevant to treatment optimization. PRS can support primary prevention by quantitating the overall burden of genetic risk factors in various subpopulations and for risk stratification. In secondary prevention, PRS may help identifying high risk individuals who warrant screening for disease or enriched samples for clinical trials. In tertiary prevention, use of single or multiple PRS in a precision medicine framework may provide criterion for decisions about optimal medications and/or lifestyle interventions tailored to genetic risk and protective factors and for the avoidance of side-effects of specific treatments.

2.7 POLYGENIC RISK SCORE IN ALZHEIMER'S DISEASE

Late onset AD (LOAD) is a highly prevalent neurodegenerative dementia characterized pathologically by brain accumulation of amyloid beta (Aβ) plaques and neurofibrillary tangles composed of hyperphosphorylated tau. These classic pathological hallmarks of AD are only the most obvious manifestation and belie a broad array of pathophysiological changes affecting numerous systems within the brain and periphery. A small percent of AD cases, typically with an early onset (EOAD) and aggressive course, are monogenic with an autosomal dominant inheritance pattern. Over 95% of AD is genetically complex, highly heritable, and therefore well-suited to polygenic investigation including analysis of heterogeneity and subgroups to support development of a precision medicine approach. Since the mechanistic drivers of LOAD remains unclear, substantial effort is being dedicated to genetic risk score modelling for individual risk prediction and to a systems approach to understanding disease pathogenesis.

APOE ε4, the strongest genetic variant associated with increased risk and earlier onset of LOAD, only partially accounts for the estimated heritability [29]. The contribution of other genetic markers has frequently been highlighted by PRS [51, 71, 76, 77]. One PRS study including 19 non-*APOE* SNPs successfully stratified *APOE ε4* carriers into risk subgroups where those with the highest scores exceeded the risk of those with the lowest score by 62% [61]. Another PRS study using 31 non-*APOE* SNPs found that age at onset (AAO) of AD is modulated by the genetic score [15]. *APOE ε3/ε3* carriers in the highest AD risk stratum, could progress to AD as many as 10 years faster than those in the lowest group [15]. Non-*APOE* PRS has also been associated with disease stage and progression (e.g., MCI-converters [34] and cognitively normal individuals [15, 36]), suggesting that genetic contributions to AD manifest in a stage-specific manner [36]. In addition, non-*APOE* PRS have been used for AD-patient classification [15, 51, 56, 61, 71, 76-80] and AD-subtype discrimination [36], which has helped to reveal diverse mechanisms underlying various AD subtypes.

In addition to clinical indicators of disease status, endophenotypes such as cerebrospinal fluid (CSF) and MRI and PET imaging measures are important AD biomarkers. In most studies, their relation to the genetic composite score was either driven by the *APOE* [38] or could not be established [35, 38, 74, 81] (possibly due to low statistical power and a small number of SNPs in the PRS [39, 75, 81]). One study [16] observed that relaxing the SNP inclusion threshold from the conventional GWAS-based $p<5x10^{-8}$ to a nominal $p<0.01$ led to several associations becoming significant, even after excluding *APOE*. This

result, however, was not replicated in other studies [36, 75]. The optimal threshold remains an open question and may be related to multiple factors as discussed above.

Accepted CSF biomarkers for AD include $A\beta_{1-42}$, total tau (t-tau), and phosphorylated tau (p-tau). However, the relation between genetic scores and these CSF biomarkers has not been consistent. Genetic association studies of $A\beta_{1-42}$ with non-*APOE* PRS were not successful in the past [71, 74]. The evidence for the PRS's relation to p-tau [71], t-tau [15, 71] and p-tau/ $A\beta_{1-42}$ ratio [77] remains limited. Recently, Desikan *et al*. [82] observed that their PHS was associated with increased intracranial Aβ plaque accumulation over time (p-value = $1.28x10^{-7}$). In another study [37], the variability explained for $A\beta_{1-42}$ was increased by 1.8%, when in addition to *APOE* other markers were included in PRS.

For neuroimaging measures, many studies have failed to detect a significant association of PRS with baseline AD imaging phenotypes (e.g., hippocampal volume) in cognitively normal older adults [81], young adults and older MCI individuals [80]. However, when older adults from 4 cohorts where combined into one large sample (>1,600 individuals), the same analysis revealed significant association of the PRS with the mean hippocampal volume at the baseline [80]. In a more recent study [15], PRS was associated with longitudinal volume loss, in both hippocampal and entorhinal cortex areas. In cognitively normal adults, a PRS was marginally associated with annual cortical thinning rates [57] and significantly associated with bi-annual hippocampal complex thinning rates [81].

Currently, PRS seems to be a useful tool for predicting the AAO of AD [15, 49, 71, 76, 77] for both sporadic late and early onset [77], even after excluding *APOE*. However, the

degree of prediction varies across studies. One unit increase in the non-*APOE* PRS is estimated to accelerate the AAO by 8 months to a year [76, 77]. Another study with >1,300 AD patients suggested that a unit increase in PRS (22 IGAP SNPs, including *APOE*) decreases the AAO by up to 2.4 years [71]. As above, *APOE ε3/ε3* homozygotes showed PRS strata differences in AAO can reach 10 years [15].

Using PRS, the maximum case/control classification accuracy level of most AD studies is ~78% [51, 71, 77, 78]. Although PRS is not sufficiently accurate for clinical classification, more important applications are subtype stratification and prediction of disease trajectory. Prediction analysis requires larger sample sizes compared to association analysis [17] but the goal of prognostic prediction may be within range. The AD heritability explained by additive genetic effects as captured by GWAS is estimated to be 24%-33% [29, 83] with the majority attributed to APOE [82]. The sample size required to observe reliable PRS effect for prediction is a function of disease heritability [17]. The largest AD GWAS [84] included 25,580 AD cases and 48,466 controls. As sample sizes continue to increase rapidly the AUC is expected to soon reach levels acceptable for clinical application. Ongoing efforts to improve the accuracy and interpretability of PRS can also be expected to advance our knowledge about AD pathogenesis and help to identify new combinatorial diagnostic/biomarker strategy for the early intervention.

2.8 CONCLUDING REMARKS

Polygenic composite score approaches have been used to identify optimized sets of SNPs whose cumulative genetic effect can better identify susceptibility and predict AAO and

phenotypic features that characterize complex diseases. With applications in a wide range of disease, PRS, the most common genetic composite score, has promise for patient screening and genetic enrichment for therapeutic intervention trials. In AD research, PRS have contributed to risk stratification for early detection and helped to elucidate the genetic contribution to disease endophenotypes.

Despite the advances in PRS methodologies discussed above, current polygenic composite score approaches have limitations, including extent of ability to account for disease heritability and insufficient development for full clinical deployment in precision medicine. A number of strategies may lead to better PRS performance (see Outstanding Questions). While current methods focus on additive effects and common variants, future approaches may incorporate the potential role of epistasis and gene-environment interactions, transcriptomic and epigenetic variation, and other patient information through combinatorial strategies. Recent advances in machine learning can be expected to improve PRS models. Another limitation is interpretability. PRS reflect enriched pathways but the downstream mechanisms through which they influence disease is not identified. New computational biology tools and databases can be expected to enhance interpretation of polygenic effects. Future polygenic models developed in relation to quantitative endophenotype data from disease specific biomarkers hold promise for clinically and mechanistically useful prediction. We can look forward to further development of these methods to support the evolving precision medicine of complex disease.

Chapter 3

ENDOPHENOTYPE-BASED POLYGENIC RISK SCORES: PREDICTION OF BIOMARKER AND CLINICAL PROGRESSION AND DEMENTIA

In its most simple version, PRS is a weighted sum of alleles frequencies, but this might not be the most efficient formulation, especially for complex diseases. In the previous chapter we went over several factors, some data-related and some methodology-related, that could have a significant on impact the performance of the score. We overviewed approaches concerning re-adjustment of the original SNP-weights, some of which incorporate biological-relevant information of the SNPs. We came across approaches that showcasing the benefits of incorporating biological-relevant information in the PRS, including their increased interpretability. By studying the existing literature on PRS exploitation in AD, we came to the realization that most effort is focused on case-control PRS refinement and application, and there is only limited evidence on how biological-relevant information could contribute to the understanding of AD-related outcomes.

3.1 INTRODUCTION

According to the multifactorial etiology [85] for complex diseases, the phenotypic variability can be explained by the additive effect of multiple genetic factors. A PRS is a mathematical formulation of this hypothesis, being a single combinatorial measure of multiple individual genetic effects that express an individual's overall genetic liability [5, 86]. In Alzheimer's disease, PRS studies have focused primarily on risk and prognosis [15, 74, 78, 86-91], with the majority focusing on late onset AD (LOAD) based on large case-

control genome wide association studies (GWAS). However, the increased phenotypic and genetic heterogeneity among LOAD patients calls for more personalized solutions and thus, for approaches that integrate biologically relevant genetic information [15, 74, 89, 92-94]. Here we employed AD endophenotype-specific GWAS to develop individual and combined endophenotype-PRSs. Our goal was to investigate the potential of endophenotype-PRSs for prediction of biomarker progression and prognosis of dementia.

3.2 MATERIALS AND METHODS

Briefly, in this work we first studied whether the individual endophenotype and combined endophenotype-PRSs can capture the risk of dementia (expressed as both odds ratio and hazard ratio) and the age of dementia diagnosis. Next, we tested the potential of endophenotype-PRS to capture the genetic risk beyond *APOE* by examining differences in dementia risk and median survival among ε3/ε3 participants. Finally, we computed linear mixed models to determine the relationship with longitudinal trajectories of known AD endophenotypes.

3.2.1 STUDY POPULATION

For the analysis, we used data from the publicly available ADNI1,GO/2 study [95]. ADNI is a multisite study that aims to track the progression of AD across the entire spectrum to discover new biomarkers, understand the relations between them, and develop treatments and optimize clinical trial measures. Since 2003, ADNI has collected clinical, imaging, genetic and biospecimen data for individuals over the age of 55 within the U.S and Canada.

For the purpose of developing endophenotype-PRSs, we focused on 11 biomarkers from ADNI1,GO/2, each biomarker representing either amyloid, tau, neuronal or vascular pathology (Figure 3.1). From the 1,550 individuals available, only 585 participants had complete baseline information on all 11 biomarkers of interest. Of these 585 participants, 80% were used for training, 20% for validation, and the remaining 965 participants were used for testing (Table 3.1) [96]. Diagnosis was based on clinical criteria and consisted of five different categories: cognitively normal (CN), significant memory concerns (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI) and demented (Dem), with dementia being characterized as participants whose diagnosis was based on clinical rather than pathological evidence [97].

Table 3.1: Data description

Characteristic	Full		Training		Validation		Testing	
Count	1550		468		117		965	
Age								
Mean	73.4		72.2		71.6		74.4	
Range	47-91		47-91		55-88		54-90	
Gender (%)								
Male	59.3%		53.5%		50%		58.8%	
Female	40.7%		46.5%		50%		41.2%	
Diagnosis (%)	Baseline Diagnosis	Final Diagnosis	Baseline Diagnosis	Final Diagnosis	Baseline Diagnosis	Final Diagnosis	Baseline Diagnosis	Final Diagnosis
CN	23.7	22.1	21.8	22.1	19.5	22.0	24.2	21.5
SMC	6.0	5.7	12.0	11.6	11.9	11.9	2.4	2.2
EMCI	18.0	17.5	29.8	26.6	32.2	26.3	10.4	11.9
LMCI	33.1	20.0	18.6	10.9	18.6	11.0	41.5	25.8
Dem	19.2	34.7	17.8	28.9	17.8	28.8	21.5	38.7

3.2.2 BIOMARKER PCA

We focused on integrating information from 11 biomarkers that fall under the A/T/N/V framework [97]. This is an expansion of the A/T/N framework [98], which was developed to reflect the pathophysiology progress of the disease and thus, provide a better understanding of its clinical stages. The set of biomarkers that we selected for PRS development is presented in Figure 3.1. These include CSF and PET amyloid (A), CSF tau

(T), MRI and FDG-PET (N) from selected regions of interest (ROI), as well as white matter hyperintensity (V). To summarize the information from these biomarkers, we performed principal components analysis (PCA) simultaneously on the residuals of all 11 biomarkers that were first pre-adjusted for age, sex, years of education, and the first two genetic PCs that controlled for population stratification. PCA was applied on the 585 individuals with full baseline biomarker data (Figure 3.1). For each participant, the baseline was defined as the first time point with available measurements for all 11 biomarkers. The analysis returned 4 components, each representing one biomarker group (A, T, N, and V).

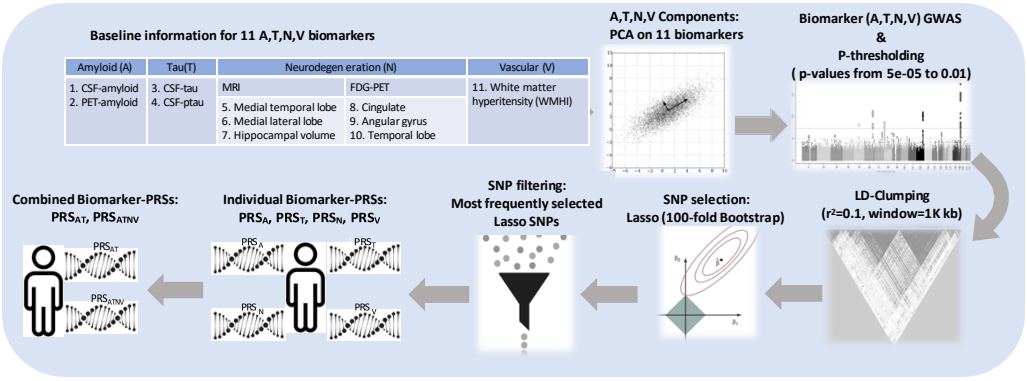

Figure 3.1: PRS calculation steps

3.2.3 SINGLE NUCLEOTIDE POLYMORPHISM (SNP) FILTERING

For each of the 4 endophenotype components obtained, we ran GWASs on the same 585 participants that had been used for the PCA step (Figure 3.1). The genotype data were HRC imputed, with a total number of 5,406,481 SNPs. The GWASs results have been filtered based on a range of p-value cut-offs (5e-05, 8e-05, 1e-04, 8e-04, 1e-03, 5e-03, 1e-02). To address the linkage disequilibrium problem (LD), we performed clumping using PLINK on SNPs with MAF≥5%, r^2=0.1 and window=1K kb. From the *APOE* region (defined as 1Mb up and downstream of the gene position, 44,409,039 to 46,412,650) only *rs429358*

and *rs7412* were retained.

3.2.4 FURTHER SNP FILTERING AND SNP WEIGHT CALCULATION

In addition to p-thresholding, we further filtered the SNPs by applying Lasso [99], a type of penalized regression. At each p-threshold and for each biomarker component, Lasso returned a list of SNPs and their corresponding weights. The Lasso penalty was determined by tuning the lambda parameter using 10-fold cross-validation. The criterion for optimal lambda selection was minimization of the mean square error (MSE). While shrinkage was applied to SNPs, the baseline age, sex, years of education, and the two *APOE* SNPs, *rs429358 and rs7412,* were not subject to penalization. To increase the stability of the result, Lasso was repeated on 100 bootstrap samples from the training set, returning at each iteration a list of SNPs and SNP weights. The final SNP list was obtained by retaining the most frequently selected SNPs (selection frequency $\geq$ 80%, Figure 3.1).

According to the literature, re-weighted SNP coefficients may achieve improved PRS performance [15, 68] compared to the traditional case/control GWAS SNP effects. Because Lasso estimates tend to be biased [100], we followed a two-step procedure by refitting a linear regression model on the Lasso selected SNPs. The regression model was adjusted for the covariates, age, sex, and years of education and was performed separately for each of the four endophenotype components. The process was bootstrapped 100 times on the training set, and the final PRS SNP weight was calculated by averaging the corresponding regression coefficient over the 100 bootstrap iterations, as described in equation (1). Here, w_s is the new weight for SNP *s*, i is the bootstrap iteration index, N is the total number of bootstrap iterations (in this case $N = 100$), and $coef_{si}$ is the regression coefficient for the SNP *s* at the i^{th} iteration.

$$w_s = \sum_{i=1}^{N} coef_{si} \Big/ N \qquad (1)$$

3.2.5 INDIVIDUAL AND COMBINED BIOMARKER-PRS

At each p-threshold and for every participant j, we calculated four individual endophenotype-PRSs (PRS_A, PRS_T, PRS_N, PRS_V) based on equation (2). The PRS was expressed as the sum over the weighted number of alleles per SNPs. Specifically, for the j^{th} individual, the endophenotype b PRS (PRS_{bj}) was obtained by multiplying the minor allele count d_{sj} of the SNP s by the SNP weight w_s (described in equation 1).

$$PRS_{bj} = \sum_{s=1}^{S_b} w_s d_{sj} \qquad (2)$$

Finally, we generated two combined endophenotype-PRSs (PRS_{ATNV}, PRS_{AT}) for each individual. The PRS_{ATNV} was expressed as the weighted sum of the individual biomarker- PRSs as shown in (3). In equation (3), PRS_b is the individual endophenotype-PRS as described in equation (2). To obtain the weights w_b, we used the training set to regress each of the four endophenotype components on the corresponding PRS_b while controlling for age, sex and years of education. The final weight w_b for each PRS_b was the average coefficient over 100 bootstrap iterations. A similar approach was followed for generating the PRS_{AT}.

$$PRS_{ATNV} = \sum\nolimits_{b \in \{A,T,N,V\}} w_b PRS_b \qquad (3)$$

3.2.6 PRS THRESHOLD SELECTION

Deciding on the GWAS p-threshold is important as it directly affects the number of SNPs to be considered in a PRS and subsequently the performance of the score. To select the optimal p-threshold and thus the final PRS for the remainder of the analysis, we assessed the prediction performance of the endophenotype-PRSs on the validation set for each of the seven p-thresholds. Specifically, we obtained the adjusted variance explained (Adj.R^2) by regressing each endophenotype on the corresponding endophenotype-PRS while controlling for baseline age, sex and years of education. The average Adj.R^2 over 100 bootstrap iterations indicated the best overall p-threshold.

3.2.7 DEMENTIA RISK IN RELATION TO PRS

To study the association between the six PRSs and the risk of dementia, we ran a logistic regression model, treating the PRS as a predictor while adjusting for the centered covariates of age, sex, and years of education. In the model described here, age was defined as either the age of clinical diagnosis of dementia or the age at the last clinical visit for the non-demented participants. To simplify the interpretation, the PRSs, originally ranging from 0 to 1 with values closer to 1 indicating higher risk, were multiplied by 10. Among the 585 participants of the training set, 367 individuals that were either CN, SMC or Dem were used for model training. Having estimated the odds of AD for each PRS, we replicated the results on 712 individuals from the testing set, after excluding MCI patients. As an additional step to assess the predictive ability of SNPs beyond *APOE*, we obtained the risk of developing dementia among $\varepsilon 3/\varepsilon 3$ participants.

3.2.8 DEMENTIA HAZARD AND AGE TO DEMENTIA DIAGNOSIS IN RELATION TO PRS

Other statistical measures of interest in AD research include the hazard of dementia and the age of dementia onset. To assess the strength of the relationship between these measures and the biomarker PRSs, we ran a Cox proportional hazard (PH) model, which was trained using 367 individuals from the training set. The "event" was considered the diagnosis of dementia (clinical manifestation), and age at dementia diagnosis (AAD) was treated as survival time in the model. PRS was used as a predictor in the model, after adjusting for the years of education and sex. To simplify interpretation, the PRSs were multiplied by 10 and education was centered. The PH assumption was tested using the cox.zph() function in R. To get predictions of the age at dementia diagnosis among the Dem cases, we predicted the survival curves using the Cox model that was previously applied on the training data. The actual and the predicted age to dementia were divided into deciles. The relationship between the predicted age and actual age of dementia onset was assessed using Pearson correlation (r). The analysis was replicated on 712 non-MCI individuals from the test set as well as on the $\varepsilon 3/\varepsilon 3$ participants.

3.2.9 PRS FOR BASELINE LEVELS AND LONGITUDINAL TRAJECTORIES OF RESPONSES OF INTEREST

In addition to dementia risk prediction, which may be useful at the prevention stage, information about disease progression and key outcomes are also important. Here, we assessed the baseline and longitudinal effects of PRSs on 14 responses of interest. These included three cognitive measures (ADNI-MEM, ADNI-EF and FAQ), as well as 11

biomarkers that were described earlier (Figure 3.1). For each of the 14 responses, we applied a random intercept linear mixed model to account for the correlation between repeated measurements. The data were aligned for all participants, with time 0 representing the first visit when a measurement was available for each biomarker. All biomarkers were transformed to range between 0 to 1, and MRI biomarkers were pre-adjusted for intracranial volume (ICV). Whenever necessary, the biomarkers were log_{10} transformed. The model was adjusted for sex and centered covariates, including years of education and baseline age. Fixed effects included years since baseline, as well as the PRS and their interaction. To simplify interpretation, the PRSs were multiplied by 10. The random intercept term allowed for varying intercepts among the participants. The performance was assessed by the Nakagawa's marginal pseudo-R^2 on the testing set. The significance of the increase in the pseudo-R^2 was assessed by ANOVA, which compared the (full) model, PRS and its interaction with time, to the (base) model, which contained covariates only. The p-values of the main PRS effect (baseline effect) and the interaction effect (longitudinal change) were corrected for multiple comparisons. Specifically, for each endophenotype, a Bonferroni correction was applied to account for testing against six PRSs (Bonferroni p-value=8.3e-03).

3.3 RESULTS

3.3.1 PRS CALCULATION

The best PRS performance was achieved for the GWAS p-value threshold of 8e-04, based on the average Adj.R^2 over 100 bootstrap iterations. Because that was the best threshold for all biomarkers, except for vascular, we considered it to be the overall optimal p-

threshold. At this optimal threshold, the number of PRS SNPs (including the *APOE* SNPs *rs429358* and *rs7412*) was 145 for PRS_A, 166 for PRS_T, 160 for PRS_N and 159 for PRS_V. The PRSs calculated at the specific threshold were used in steps for the remaining analysis.

3.3.2 DEMENTIA RISK IN RELATION TO PRS

On the training set, the strongest association between clinically diagnosed dementia and PRS was observed for PRS_N followed by $PRS_{ATNV.}$ For the former, a 0.1 unit increase in the PRS_N increased the odds of dementia by 4.5 times (OR=4.5, *P*=1.28e-20), whereas for the latter, the OR of Dem was 3.38 (*P* =9.96e-24). The results were validated on the testing set including all *APOE* groups (PRS_N: OR=1.29, *P* =4.8e-04; PRS_{ATNV}: OR=1.52, *P*=1.03e-07). To demonstrate the information provided by the SNPs beyond *APOE*, we examined the strength of the association among $\varepsilon3/\varepsilon3$ carriers and observed a significant OR of 4.7 (*P* =1.45e-08) for PRS_N and 2.76 for PRS_{ATNV} (*P* =7.58e-10) on the training set. On the $63/63$ testing group, neither PRS_N nor PRS_{ATNV} effects were significant (PRS_N: OR=1.13, *P* >0.1; PRS_{ATNV}: OR=1.19, *P* >0.1).

3.3.3 DEMENTIA HAZARD AND AGE TO DEMENTIA DIAGNOSIS IN RELATION TO PRS

On the training set, the strongest association between dementia onset and PRS was observed for PRS_{AT} followed by $PRS_{ATNV.}$ For the former, the rate of being clinically diagnosed with dementia at any time point was increased by 67% for each 0.1 unit increase of the PRS_{AT} (*P* =3.52e-26), whereas for the latter, the hazard ratio (HR) of AD was 1.62 (*P* =2.04e-25). Both associations were replicated on the test set (PRS_{AT}: HR=1.24,

P=5.97e-07; PRS_{ATNV}: HR=1.20, *P*=2.08e-05). Among $\varepsilon 3/\varepsilon 3$ carriers of the training set, both PRS effects were significant (PRS_{AT}: HR=1.53, *P*=4.74e-06; PRS_{ATNV}: HR=1.58, *P*=3.49e-08). We additionally obtained a 10-year difference in the median AAO between the extreme PRS_{AT} quartiles ($PRS_{AT,Q1} \leq 0.29$, $PRS_{AT,Q4} \geq 0.60$) of the $\varepsilon 3/\varepsilon 3$ in the training set (Figure 3.2A; AAD: 76 for Q4 and 86 for Q1).

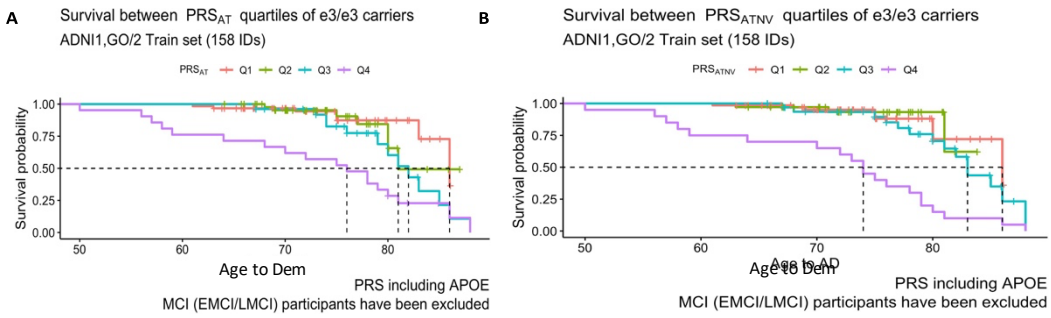

Figure 3.2: Survival curves among ε3/ε3 individuals in the training set. Dashed lines represent the median age to dementia. (A) Results by PRS_{AT} quartiles. (B) Results by PRS_{ATNV} quartiles.

For PRS_{ATNV} the median AAD was 86 in Q1 and 74 in Q4 ($PRS_{ATNV,\ Q1} \leq 0.33$, $PRS_{ATNV,\ Q4} \geq 0.63$) (Figure 3.2B). The significance of the AAD prediction was not replicated in the test set. For all the results reported here, the proportional hazard (PH) assumption was met. Finally, we evaluated the scores' performance in predicting the AAD by observing its association with the actual AAD. The association was observed among deciles of the predicted and observed AAD. For the training set, the Pearson correlations were r_{AT} = 0.83 (*P*=2.7e-03) and r_{ATNV} = 0.78 (*P*=7.4e-03). For the testing set, the correlations were r_{AT} = 0.76 (*P*=1.2e-02) and r_{ATNV} = 0.64 (*P*=4.6e-02).

3.3.4 PRS AND LONGITUDINAL TRAJECTORIES OF COGNITIONAL AND BIOMARKER RESPONSES

We compared the percentage variance explained for the mixed model with and without PRS and assessed the change using ANOVA to test between the two models. The results on the test set are presented in Table 3.2.

Table 3.2: Marginal Variance explained (R^2) increase due to endophenotype-PRS. Longitudinal analysis results using PRS that includes *APOE*. Results on the ADNI1,GO/2 test set.

Endophenotype	Base model	PRS_A	PRS_T	PRS_N	PRS_V	PRS_{AT}	PRS_{ATNV}
ATNV-related endophenotypes							
Roche_CSF_ABETA	0.93%	**9.22%**	1.98%	1.12%	--	8.82%	7.96%
AV45_WC_SUM	0.70%	11.25%	4.08%	2.48%	0.58%	12.38%	**12.76%**
Roche_CSF_TAU	2.06%	1.55%	**6.37%**	--	0.09%	5.47%	4.50%
Roche_CSF_PTAU	1.23%	2.40%	**7.10%**	--	0.07%	7.00%	5.41%
MeanLatTemp_ThxAvg	11.19%	0.26%	0.68%	0.74%	0.09%	0.69%	**1.22%**
MeanMedTemp_noLingual_ThxAvg	12.30%	0.77%	1.29%	1.18%	0.11%	1.57%	**1.78%**
Hippocampus_Vol	13.72%	1.86%	1.48%	0.52%	0.20%	**2.78%**	2.17%
FDG_TempLobe	3.10%	1.65%	2.82%	1.11%	--	3.51%	**4.20%**
FDG_AngGyrus	3.62%	1.41%	2.21%	0.82%	--	2.88%	**3.58%**
FDG_Cingulate	6.00%	1.41%	1.74%	1.19%	0.54%	2.57%	**3.90%**
WMHI	5.84%	0.30%	--	0.13%	**0.40%**	0.28%	**0.40%**
Not ATNV-related endophenotypes	**Base model**	**PRS_A**	**PRS_T**	**PRS_N**	**PRS_V**	**PRS_{AT}**	**PRS_{ATNV}**
ADNI_MEM	8.14%	1.71%	--	1.08%	--	1.79%	**2.29%**
ADNI_EF	8.62%	--	**0.65%**	--	--	0.40%	0.51%
FAQ	4.52%	1.12%	1.35%	0.95%	--	2.02%	**2.28%**

Base mixed model $Y = \beta_0 + \beta_1 t + covar + (1|ID)$

Full mixed model $Y = \beta_0 + \beta_1 t + \beta_2 PRS + \beta_3 (PRS*t) + covar + (1|ID)$

Overall increase represents the increase caused due to inclusion of PRS and its interaction with time all together

Dashes indicate that the overall R^2 increase compared to the base model (covariates ony) was not significant (ANOVA p-value> 0.05)

For CSF- amyloid, tau, and p-tau, the individual endophenotype-PRSs for amyloid, and tau (PRS_A, PRS_T) resulted in the highest improvement of the explained variance for the corresponding biomarkers. This improvement was statistically significant after correcting for multiple comparisons (Table 3.2). By examining the significance of the PRS terms in these models, we noticed significant main effects for the PRS, but the interaction terms were insignificant for all three biomarkers (Table 3.3). Significant interaction for both tau biomarkers was achieved by PRS_{ATNV}, although this was not the optimal score in terms of variance explained (Table 3.2, CSF-tau: $R^2_{PRS_{ATNV}}$= 4.50%; CSF-ptau: $R^2_{PRS_{ATNV}}$= 5.41%). When studied separately, PRS_A and PRS_V showed a negative relation to both tau measures

(CSF-tau: Time×PRS_A: *P*=6.5e-03; Time×PRS_V: *P*=5.5e-03; CSF-ptau: Time×PRS_V: *P*=2.3e-03; Time×PRS_{ATNV}: *P*=5.9e-02). On the other hand, PET-amyloid, MRI and FDG--PET biomarkers, as well as memory and FAQ had stronger associations to combined- PRSs (Table 3.2). Exception to that were PET-amyloid and MEM, where levels were strongly linked to the combined PRS levels both at the baseline and longitudinally (Table 3.3), even after correcting for multiple comparisons (Table 3.3).

Table 3.3: *P*-values for endophenotype-PRS and its interaction with time. Longitudinal analysis results using PRS that includes *APOE*. Results on the ADNI1,GO/2 test set.

Endophenotype	**With APOE**		
ATNV-related endophenotypes	**Best PRS**	**Main Effect (β_2)**	**Interaction (β_3)**
Roche_CSF_ABETA	A	-0.04**	--
AV45_WC_SUM	ATNV	0.05**	--
Roche_CSF_TAU	T	0.04**	--
Roche_CSF_PTAU	T	0.05**	--
MeanLatTemp_ThxAvg	ATNV	-0.01**	-1.8e-03**
MeanMedTemp_noLingual_ThxAvg	ATNV	-0.02**	-3.3e-03**
Hippocampus_Vol	AT	-0.01**	-1.2e-03**
FDG_TempLobe	ATNV	-0.03**	-4.1e-03**
FDG_AngGyrus	ATNV	-0.02**	-2.4e-03**
FDG_Cingulate	ATNV	-0.02**	-2.7e-03**
WMHI	V	0.01*	--
Not ATNV-related endophenotypes	**Best PRS**	**Main Effect (β_2)**	**Interaction (β_3)**
ADNI_MEM	ATNV	-0.02**	--
ADNI_EF	T	-0.01*	--
FAQ	ATNV	0.05**	4.1e-03**

Mixed model $Y = \beta_0 + \beta_1 t + \beta_2 PRS + \beta_3 (PRS*t) + covar + (1|ID)$ Dashes indicate insignificant result (p > 0.05)

** p< 8.3e-03 (Bonferroni correction by biomarker p-value: 0.05/6=8.3e-03)

* p< 0.05

Overall, most of the associations studied here remained significant within the ε3/ε3 training set, but none of the neurodegeneration markers reached significance on the ε3/ε3 testing set. FAQ was the only cognition measure that remained significant, even among the ε3/ε3 individuals of the test set.

3.4 DISCUSSION

We developed individual and combined endophenotype-PRSs and evaluated their association with dementia risk, age at dementia diagnosis and biomarker trajectories. Combined endophenotype-PRSs led to significantly higher dementia hazard and specifically accelerated the median AAO among ε3/ε3 participants up to 12 years. Finally, PRS_A and PRS_T, which were AD-specific, were better predictors of amyloid and tau biomarkers, while combined endophenotype-PRSs were better predictors of neurodegeneration.

We showed that PRSs based on specific AD biomarkers can be used to assess dementia risk and prediction of biomarker trajectory [101]. In addition, we found that the progression of pathophysiological biomarkers and cognitive decline have a stronger association with the combined-PRSs, compared to the more biologically restricted individual endophenotype-PRSs. A possible explanation is that the combined-PRSs accounts for the effect of SNPs related to multiple endophenotypes and thus, better captures the multiple biological mechanisms implicated in these biomarkers. The insignificant interaction effects of time with PRS_A and PRS_T, when modeling the CSF amyloid and tau trajectories respectively (Table 3.3), may indicate that the observed increase in explained variance was likely driven by the strong association of these scores with the corresponding baseline biomarker levels. One the other hand, the significant $PRS_{ATNV} \times \text{Time}$ interaction for CSF-tau and CSF-ptau may emanate from the amyloid and vascular terms integrated in PRS_{ATNV}, which seem to have a negative association to both tau trajectories. This could support the idea that different PRSs may be preferred, depending on whether we are

interested in predicting cross-sectional differences or differences in the rate of change. Lastly, we provided significant evidence for genetic risk beyond *APOE* by replicating the previously observed differences in the age of dementia onset among ε3/ε3 participants [15].

In this study, we found that PRS accounts for biologically relevant information that may elucidate the level of genetic complexity of AD endophenotypes and related outcomes. Superiority in performance of combined-PRSs compared to individual-PRSs may indicate greater complexity in the underlying biological mechanisms of the response of interest, which may be indicative of the additional genetic information incorporated in the combined-PRS. While the intention in generating endophenotype-PRSs was partially the improvement of PRS's interpretability, other scores with the same goal have been developed. Pathway-PRS is one such score, which attempts to increase interpretability by inclusion of SNPs that are part of a specific biological pathway [74, 92, 101]. Despite the seemingly similar rationale between the proposed individual endophenotype-PRS and pathway-PRS [74, 92, 101], there are also major differences. For example, the existing pathway-PRSs are developed based on case/control GWAS that may fail to identify SNPs related to important biomarkers [102]. This might be especially plausible when the disease endophenotypes are closer to the molecular mechanism than the disease status, in which case endophenotype-GWASs may have greater utility in identifying biomarker-related SNPs [102]. Pathway-PRS also requires *apriori* knowledge about the disease pathways and the SNPs belonging to that pathway, which may restrict the number of informative SNPs in the PRS and thereby the results of the analysis [5]. In contrast, the endophenotype-PRS

identifies biomarker-related SNPs through endophenotype-GWAS, allowing multiple biological pathways associated with that biomarker to enter the score at once.

In this work, we provided a comprehensive comparison between individual and combined endophenotype-PRSs based on *ADNI* data. However, our work is not free of limitations. First, there was a limited sample size for PRS development, as *ADNI* is the only publicly available study that offers such an extensive collection of AD-related biomarkers. Limited sample size is also a barrier because the data had to be further split into training and testing conditions. However, as more participants are recruited, the power of the analysis will improve. The limited discovery sample may also explain our failure to observe a significant AAO difference among the ε3/ε3 participants in the testing set. Second, although part of the ADNI1,GO2 was kept separate and used solely for replication, it is still included in the same cohort that was used for PRS development. Replication of these results in completely independent data sets is necessary and should be pursued when data availability permit. Third, *ADNI* is not ideal for building vascular-PRSs because individuals with more severe cerebrovascular disease are typically excluded. A better vascular-PRS should be derived using a more appropriate data set enriched for cerebrovascular disease.

To conclude, our study suggests that PRS_A and PRS_T are AD-specific as they have the best performance in predicting amyloid and tau biomarkers, whereas the combined PRSs are more general and preferred for predicting neurodegeneration. Also, further analysis of the endophenotype-PRSs could offer functional insights and promote treatment development. Specifically, in the context of precision medicine, endophenotype-PRSs could be used for

generating more specific genetic risk profiles for prospective trial enrollees that are specifically aligned with measurable biomarkers thought to reflect disease status. In the future, individual endophenotype scores could be utilized to study the genetic heterogeneity among individuals at risk for dementia, which could potentially provide useful information about the observed variability in the disease's clinical manifestation and further support the necessity for individualized treatment.

Chapter 4

INVESTIGATING THE LINK BETWEEN CANCER-RELATED COGNITIVE OUTCOMES AND ALZHEIMER'S PATHWAY POLYGENIC RISK SCORES AMONG OLDER BREAST CANCER SURVIVORS

We further extended the concept of biologically informed PRS, by utilizing pathway-level information of genome-wide significant AD SNPs to predict cancer-induced cognitive changes. An increasing number of studies indicate a potential link between the two diseases with cancer survivors developing cognitive changes and with biological pathways and genetic markers shared by both diseases. But overall, the findings about this relation are contradicting with some studies identifying a positive relation between the risk of the two diseases and some a negative relation. To the best of my knowledge, some work has been done studying the relation of cancer-related-cognitive impairment and neurodegeneration-related genes, but this has only been explored on individual gene level. Here we expand this effort by testing the association between breast cancer-related cognitive changes and the aggregated genetic effect of AD-related markers. AD related genetic profile could further support BC treatment allocation decisions by evaluating a patient's risk for high cognitive deficits.

4.1 INTRODUCTION

Cancer-related cognitive impairment is a common phenomenon among BC survivors [103]. The advancing cancer treatments and the fact that cancer is a disease which mainly affects older populations, will result in an increasing number of survivors whose cognitive

aging will be affected by cancer and/or cancer-treatment. Studies have shown that the duration and rate of the cognitive decline varies by systemic treatment [103-105] and potential models for this decline have been proposed [106]. Another factor that could potentially contribute to cancer-related cognitive deficiency is AD. Currently, the potential link between cancer and AD is obscure [104, 107] with contradicting findings regarding the directionality of this relationship [108-110]. Several shared pathways have been identified between cancer and AD with some being inversely regulated in the two diseases, which could support the protective relationship that some studies report [109, 111]. However, the mechanistic underpinnings remain unclear [109] and causal inference is not trivial as several factors may bias the study results [104]. Genetic overlap between AD and caner may help disentangle this link [104]. Several genetic polymorphisms have been associated both with cognitive decline and cancer, including *APOE* [106, 112]. *APOE* is the strongest genetic risk factor for AD and seems to affect the cancer-related cognitive function as well [105, 113, 114]. Older populations are more vulnerable to cognitive declines, since aging is a well-established risk factor of cognitive decline and *APOE* effect seems to be age-dependent [115-117]. Thus, is important to focus on older survivors and controls to assess the effect of age [105]. Here we extended a previous work [105] regarding the *APOE* effect on the cancer-related-cognition-impairment among *TLC* participants, by accounting for multiple AD-related genetic factors simultaneously through AD-specific pathway polygenic risk scores (PRSs) [5]. For improved interpretability of the findings and to promote the understanding of the shared biological mechanisms, we used pathway-PRSs [92]. In the past associations between pathway-PRSs and other dementia-related outcomes, such as dementia risk, progression, AD neuroimaging biomarkers [92],

and AD resilience [118] have been studied, but to the best of our knowledge, the relation to specific cognitive domains has never been explored.

4.2 METHODS

4.2.1 POPULATION

The data used for this analysis are from the *TLC* study described in [105]. Briefly, *TLC* is a prospective, multisite study designated to collect information regarding the cancer-related cognition changes in older nonmetastatic breast cancer (BC) survivors. The survivors and their matched controls were individuals 60 years of age or older, without stroke history or head injury, and free of psychiatric or neurodegenerative disorders. Matching criteria include age, race, education, and site. The participants have undergone several neuropsychological tests and biospecimens collection for genotyping. At the time of the analysis, cognitive data were available for 1,285 individuals (707 survivors and 578 controls) whereas 819 individuals had been genotyped (Figure 4.1). White non-Hispanic (WNH), genotyped individuals receiving either hormonal therapy, chemotherapy, or a combination of the two (n=726), were considered for further analysis.

4.2.2 GENOTYPING

APOE *APOE* genotyping was performed using extracted DNA for all subjects from the study with available samples. SNPs *rs7412* and *rs429358* were assessed either using TaqMan assays (Life Technologies, Carlsbad, CA) and/or Fluidigm genotyping using a

custom-designed 96-SNP genotype chip (Fluidigm, San Francisco, CA).

GWAS Genotyping for TLC subjects was performed in batches on extracted DNA with either the Affymetrix Axiom Precision Medicine Research Array (Thermo Fisher Scientific, Waltham, MA), or the Illumina Global Screening Array 2.0 (Illumina, San Diego, CA). Data processing and quality control was performed using Illumina GenomeStudio Software and Plink v1.9 [119, 120]. Quality control included verification of female genetic sex, variant call rate >95%, sample call rate >90%, and Hardy-Weinberg equilibrium (HWE) $P<1x10^{-6}$. Samples were assessed for genetic ancestry in Plink and individuals with non-white European ancestry were removed. Genotype data was imputed with the Michigan Imputation Server [121], with the Haplotype Reference Consortium (HRC) reference panel [122]. 815 participants with data passing all quality control remained after imputation. Related individuals were identified using identity-by-descent in Plink (pi-hat >0.25) and one of each pair was randomly removed prior to analyses.

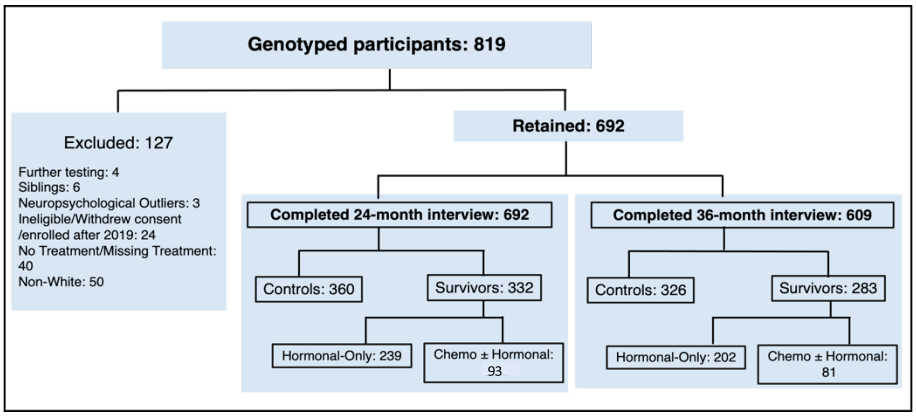

Figure 4.1: Data description

4.2.3 MEASURES

4.2.3.1 Outcomes

The main responses of interest were six composite domains of subjective cognition. These domains were: attention-processing speed-and-executive function (APE), learning and memory (LM), memory (MEM), executive function (EF), language, and visuospatial. Details regarding the neuropsychological test assigned to each of these domains are presented in Table 4.1. The development of APE and LM have been described in detail elsewhere [105]. The remaining three scores (except from the visuospatial) are composite scores harmonized across five different studies (ACT, ADNI, MAP, ROS and BLSA). Each of the six scores has been previously standardized according to the mean and standard deviation of the control group on the entire *TLC* data.

Table 4.1: Description of composite cognitive domains

	Cognitive Domain					
Number of Neuropsychological Test	**APE**	**EF[1]**	**Language[1]**	**LM**	**MEM[1]**	**Visuospatial**
1	Trail Making A	Trail Making A	Letter A fluency	Logical Memory I	Logical Memory Delayed A	Copy Score
2	Trail Making B	Trail Making B	Letter F fluency	Logical Memory II	Logical Memory Immediate	Figure Drawing Planning
3	NAB Digits Forward	NAB Digits Forward	Letter S fluency	NAB List A Immediate Recall	NAB List A Immediate Recall	
4	NAB Digits Backward	NAB Digits Backward	BNT 30 odd items	NAB List A Short Delay Recall	NAB List B Immediate Recall	
5	Digit Symbol Test	Digit Symbol Test	Animal category fluency	NAB Long Delay	NAB Long Delay	
6	COWAT	NAB Driving Scenes			List A Short Delay Recall	
7					List Learning Trial 1	
8					List Learning Trial 2	
9					List Learning Trial 3	
10					list learning long delayed forced choice recognition measure	

[1] These scores are co-calibrated/harmonized cognitive domains across the following studies: Adult Changes in Thought (ACT) study, the Alzheimer's Disease Neuroimaging Initiative (ADNI), the Rush Memory and Aging Project (MAP) and Religious Orders Study (ROS), and the Baltimore Longitudinal Study of Aging (BLSA)

4.2.3.2 Variables

The main predictors of interest were the pathway-specific PRSs and the treatment groups (3-levels: controls, hormonal-only, chemotherapy±hormonal). We examined the effect of the derived PRSs on the longitudinal and cross-sectional treatment group differences in the seven cognitive domains. The effect of PRS was tested with and without *APOE* in the score. To control for potential confounding effects, the results were adjusted for age, Wide Range Achievement Test 4 (WRAT4) and site.

4.2.4 STATISTICAL ANALYSIS

We investigated the sample characteristics (Table 4.3) such as baseline age, years of education, baseline WRAT4 score as well as the AD-related genetic risk, by calculating the corresponding mean and standard deviation. The group similarity for these variables was assessed by applying two-sample (unpaired) Wilcoxon test, as none of the variables is following Gaussian distribution (tested using Shapiro-Wilk test). The association between the *APOE* ε4 positivity (the strongest genetic risk for LOAD) with the different groups in the sample, was assessed based on the chi-square test. Linear mixed models with random intercept tested the interactions of group-by-time and group-by-time-by-PRS for each of the seven cognitive domains, treating time as categorical variable and PRS as continuous variable. To study the impact of different PRS quartiles in the cognitive changes, the analysis was repeated treating PRS as categorical variable with four levels (Q1-Q4). All models controlled for the potential confounding effect of age, WRAT4 and site which were treated as fixed variables.

4.2.5 CALCULATION OF POLYGENIC RISK SCORE

The Pathway-PRSs and the full-PRS used in this study, have been previously described elsewhere [92]. Briefly, the top 20 SNPs that have been linked to LOAD based on multiple genome wide association studies [82, 93, 123-127], were considered for construction of PRSs (Table 4.2). Among these 20 SNPs, 13 have been mapped into 8 different biological pathways with several SNPs participating in more than one pathway [128, 129]. From these, *rs75932628* in *TREM2* was removed due to low minor allele frequency (MAF). The remaining 19 SNPs have been used for the calculation of the full-PRS. Because at least two SNPs were required for PRS calculation, *protein ubiquitination* pathway was excluded from the analysis, as it was consisted by only one SNP. Since *APOE* is the strongest risk factor for AD, for the pathways containing *APOE*, we repeated the analysis with and without *rs429358* in the PRS. Thus, *hemostasis* and *hematopoietic cell lineage* pathways which compromised by only two SNPs (including *rs429358*), were considered only for the first

Table 4.2: Biological pathways used for pathway-PRS calculation

					Genes reported in pathway					
Pathway	Gene	Assigned SNP	Chromosome	IGAP SNP effect [3]	Jones et al	Guerreiro et al	Pathway-PRS Including APOE	Pathway-PRS Excluding APOE	Pathway-PRS used in current analysis	SNP included in full-PRS
Immune Response	CLU	rs9331896	8	0.146	Yes	Yes	No	Yes	Yes	Yes
	CR1	rs6656401	1	-0.157	Yes	Yes				Yes
	INPP5D	rs35349669	2	0.066	Yes	Yes				Yes
	EPHA1	rs11771145	7	-0.102	-	Yes				Yes
	MS4A6A	rs983392	11	-0.108	-	Yes				Yes
	TREM2	rs75932628	6	0.889	-	Yes				No [2]
	MEF2C	rs190982	5	0.08	-	Yes				Yes
Endocytosis	CD2AP	rs10948363	6	0.098		Yes	No	Yes	Yes	Yes
	PICALM	rs10792832	11	0.13	Yes	Yes				Yes
	BIN1	rs6733839	2	0.188	Yes	Yes				Yes
	SORL1	rs11218343	11	-0.27	-	Yes				Yes
Cholesterol transport	CLU	rs9331896	8	0.146	Yes	Yes	Yes	Yes	Yes	Yes
	ABCA7	rs4147929			Yes	Yes				Yes
	SORL1	rs11218343	11	-0.27	-	Yes				Yes
	APOE_e4	rs429358	19	1.3503	Yes	Yes				Yes
Hematopoietic cell lineage	CR1	rs6656401	1	-0.157	Yes	-	Yes	No [1]	Yes	Yes
	APOE_e4	rs429358	19	1.3503	Yes	-				Yes
Protein ubiquitination	CLU	rs9331896	8	0.146	Yes	-	No [1]	No [1]	No [1]	Yes
Hemostasis	CLU	rs9331896	8	0.146	Yes	-	No	Yes	Yes	Yes
	INPP5D	rs35349669	2	0.066	Yes	-				Yes
Clathrin AP2 adaptor complex	CLU	rs9331896	8	0.146	Yes	-	Yes	Yes	Yes	Yes
	PICALM	rs10792832	11	0.13	Yes	-				Yes
	APOE_e4	rs429358	19	1.3503	Yes	-				Yes
Protein folding	CLU	rs9331896	8	0.146	Yes	-	Yes	No [1]	Yes	Yes
	APOE_e4	rs429358	11	1.3503	Yes	-				Yes

[1] Pathways with only one SNP were not considered for analysis
[2] Removed during QC as rare variant
[3] Effect expressed as log OR

Part of the analysis. The PRS calculation was based on the logarithmic transformation of the SNPs' GWAS effects [124]. The aggregated genetic risk associated to each pathway was expressed as a weighted sum of the allele count of the corresponding pathway-SNPs.

4.3 RESULTS

Among the eligible 692 participants, the average age is 68, with a range of 60-98. Years of education varies from 6 years to 18 years with an average education of roughly 15.5 years. As shown in Table 4.2, the baseline characteristics including demographics, cognition, memory, and genetic risk were very similar ($P>0.05$) between controls and survivors. Statistically significant differences in baseline age were observed between treatment groups, with the hormonal-only groups being on average almost 2 years older than the combination therapy group (Table 4.3). Survivors in the chemotherapy±hormonal had statistically higher EF levels compared to the hormonal-only group (Table 4.3). Except from a trend for the $PRS_{(Endocytosis)}$ ($P<0.1$) no other PRS showed significant group differences (Table 4.3).

4.3.1 APE

For controls, analysis of 24-month data revealed significant APE increase over time, consistent with expected practice effects (Table 4.4). The overall group-by-time interaction was not significant (Table 4.4), indicating a similar rate of change for all groups during the 2-year interval. model, we observed a significant group-by-time-by-$PRS_{(Immune\ Response)}$ interaction (Table 4.4), indicating that the immune-related risk levels affect the rate of

cognitive changes of the treatment groups. To further examine the observed relation, we repeated the analysis using the $PRS_{(Immune\ Response)}$ quartiles as predictor, which revealed a

Table 4.3: Baseline characteristics and PRS levels

24-month follow-up								
	Controls	Survivors	*P*-value[(a)]	Total[(c)]	Hormonal Only	Chemotherapy ± Hormonal	*P*-value[(b)]	Total[(d)]
Count (%)	360(52.0%)	332(48.0%)		692(100%)	239(72.0%)	93(28.0%)		332(100%)
Baseline Age Mean(SD)	68.1(6.9)	68.0(5.8)			68.6(5.9)	66.6(5.1)	**4.3e-03	
Education Mean(SD)	15.6(2.3)	15.4(2.1)			15.5(2.1)	15.4(2.2)		
WRAT4 Mean(SD)	113.4(15.3)	112.0(15.2)			111.9(15.7)	112.2(13.6)		
APOE ε4+ Count (%)	87(24.2%)	75(22.6%)		162	58(24.3%)	17(18.3%)		75
Family history of Dementia Count (%)	134(54.9%)	110(45.1%)		244	77(32.2%)	33(35.5%)		110
EF Mean(SD)	1.65(0.34)	1.61(0.33)			1.59(0.33)	1.67(0.33)	**3.5e-02	
Language Domain Mean(SD)	0.86(0.38)	0.84(0.36)			0.84(0.35)	0.85(0.37)		
Memory Domain Mean(SD)	0.78(0.38)	0.77(0.38)			0.77(0.39)	0.79(0.37)		
Visuospatial Mean(SD)	0.09(0.80)	0.13(0.71)			0.13(0.75)	0.13(0.61)		
APE Mean(SD)	0.11(0.60)	0.01(0.62)	*3.7e-02		-0.01(0.61)	0.06(0.65)		
LM Mean(SD)	0.04(0.80)	0.06(0.78)			0.05(0.77)	0.09(0.80)		
Pathway-PRS including APOE Mean(SD)								
Cholesterol transport	0.27(0.67)	0.29(0.68)			0.31(0.70)	0.22(0.65)		
Clathrin AP2 adaptor complex	0.68(0.66)	0.69(0.68)			0.72(0.69)	0.63(0.65)		
Endocytosis	0.33(0.19)	0.31(0.18)	*8.1e-02		0.31(0.19)	0.33(0.18)		
Hematopoietic cell lineage	0.07(0.66)	0.11(0.67)			0.14(0.69)	0.03(0.62)		
Hemostasis	0.24(0.11)	0.23(0.11)			0.23(0.11)	0.23(0.11)		
Immune Response	-0.08(0.18)	-0.09(0.18)			-0.09(0.17)	-0.08(0.20)		
Protein Folding	0.52(0.66)	0.54(0.68)			0.55(0.70)	0.48(0.68)		
Full	0.30(0.74)	0.33(0.74)			0.36(0.76)	0.28(0.69)		
36-month follow-up								
	Controls	Survivors	*P*-value[(a)]	Total[(c)]	Hormonal Only	Chemotherapy ± Hormonal	*P*-value[(b)]	Total[(d)]
	326(53.5%)	283(46.5%)		609(100%)	202(71.4%)	81(28.6%)		283(100%)
Baseline Age Mean(SD)	67.8(6.9)	68.0(5.9)			68.7(6.0)	66.4(5.2)	**2.2e-03	
Education Mean(SD)	15.6(2.3)	15.3(2.1)			15.4(2.1)	15.3(2.2)		
WRAT4 Mean(SD)	113.1(15.5)	111.8(15.4)			111.8(13.1)	111.7(16.3)		
APOE ε4+ Count (%)	78(23.9%)	66(23.3%)		144	51(25.2%)	15(18.5%)		66
Family history of Dementia Count (%)	121(37.1%)	86(30.4%)	*9.6e-02	207	59(29.2%)	27(33.3%)		86
EF Mean(SD)	1.65(0.34)	1.60(0.34)	*8.0e-02		1.57(0.34)	1.68(0.34)	**1.5e-02	
Language Domain Mean(SD)	0.86(0.38)	0.83(0.37)			0.83(0.36)	0.85(0.38)		
Memory Domain Mean(SD)	0.77(0.38)	0.78(0.39)			0.77(0.40)	0.80(0.38)		
Visuospatial Mean(SD)	0.10(0.82)	0.13(0.73)			0.13(0.77)	0.15(0.62)		
APE Mean(SD)	0.10(0.61)	-2.5e-03(0.65)	*6.1e-02		-0.03(0.63)	0.06(0.68)		
LM Mean(SD)	0.03(0.80)	0.07(0.79)			0.07(0.78)	0.09(0.82)		
Pathway-PRS including APOE Mean(SD)								
Cholesterol transport	0.27(0.68)	0.29(0.69)			0.30(0.70)	0.24(0.68)		
Clathrin AP2 adaptor complex	0.68(0.67)	0.70(0.69)			0.73(0.70)	0.65(0.68)		
Endocytosis	0.34(0.19)	0.31(0.18)	*6.1e-02		0.30(0.19)	0.34(0.17)		
Hematopoietic cell lineage	0.08(0.66)	0.11(0.68)			0.14(0.69)	0.06(0.64)		
Hemostasis	0.24(0.11)	0.23(0.11)			0.23(0.11)	0.23(0.11)		
Immune Response	-0.08(0.18)	-0.08(0.17)			-0.09(0.19)	-0.08(0.17)		
Protein Folding	0.52(0.66)	0.54(0.68)			0.57(0.69)	0.49(0.67)		
Full	0.30(0.75)	0.35(0.75)			0.36(0.76)	0.33(0.70)		

(a) Test *p-value* comparing survivors versus controls.
(b) Test *p-value* comparing Chemotherapy ± Hormonal versus Hormonal-Only
(c) Total count of variable in survivors and controls
(d) Total count Chemotherapy ± Hormonal and Hormonal-Only Continuous variables were tested using either t-test or Wilcoxon test Categorical variables were tested using chi-square test

significant group-by- time-by-$PRS_{(Immune\ Response)}$ interaction (P=2.6e-02). Controls and hormonal-only participants with low $PRS_{(Immune\ Response)}$ risk, exhibited the expected

practice effects by achieving a significant improvement in their performance at the end of the second year (Table 4.5, Figure 4.2). In contrast, survivors with low genetic risk that followed combined therapy did not only fail to achieve practice effect but had a decline

Table 4.4: Significance of cognitive changes by treatment group, *APOE* and PRS

		24 months dataset (N=692)						**36 months dataset (N=609)**					
		Overall *P*-value						Overall *P*-value					
		APE	LM	EF	MEM	Language	Visuospatial	APE	LM	EF	MEM	Language	Visuospatial
	Time (months)	3.2e-05		2.9e-06	2.10e-10	3.8e-03	4.0e-02	6.7e-05	1.1e-15	1.4e-05	1.2e-13	6.2e-03	
	Group-by-Time							5.5e-01					
	Group-by-Time-by-*APOE*												
PRS Type With *APOE*	Group-by-Time-by-*PRS*												
	Overall												
	Cholesterol transport						4.2e-02						1.2e-02
	Clathrin AP2 adaptor complex						3.3e-02						1.3e-02
	Endocytosis				3.7e-02							5.5e-02	
	Hematopoietic cell lineage						3.2e-02						
	Hemostasis												
	Immune Response	8.9e-03		5.3e-02			3.7e-02	6.6e-03		5.5e-02			1.7e-02
	Protein Folding						3.8e-02						9.9e-03
PRS Type Without *APOE*	Group-by-Time-by-*PRS*												
	Overall												
	Cholesterol transport												
	Clathrin AP2 adaptor complex												
	Endocytosis				3.7e-02							5.5e-02	
	Hematopoietic cell lineage	--	--	--	--	--	--	--	--	--	--	--	--
	Hemostasis												
	Immune Response	8.9e-03		5.3e-02			3.7e-02	6.6e-03		5.5e-02			1.7e-02
	Protein Folding	--	--	--	--	--	--	--	--	--	--	--	--

Time and Treatment groups have been treated as categorical variables, whereas pathway-PRS has been treated as continuous variable. Models adjust for age, WRAT4 and site which were treated as fixed effects.
Models include up to 3-way interactions along with all lower degree interactions Empty cells indicate *P* >5.5e-02
Dashes indicate that the pathway contains only one SNP and was not used in the analysis

in their overall performance, although not statistically significant (Table 4.5, Figure 4.2). None the remaining pathway-PRSs nor the full-PRS had significant group-by-time interaction (Table 4.4). The significance of the group-by-time-by-$PRS_{(\text{Immune Response})}$ interaction was also observed on the 36-month dataset (Table 4.4).

4.3.2 LM

During the first two years, all groups exhibited practice effects for LM, with final scores that were significantly improved compared to the baseline (P<8.9e-03). The overall group-by-time interaction as well as the group-by-time-by-*APOE* were insignificant, indicating similar improvement rates for all groups that were not affected by the participants' *APOE ε4* status (Table 4.4). The only significant difference we observed at month 12 where *ε4+* survivors in the hormonal-only group had on average significantly lower LM levels

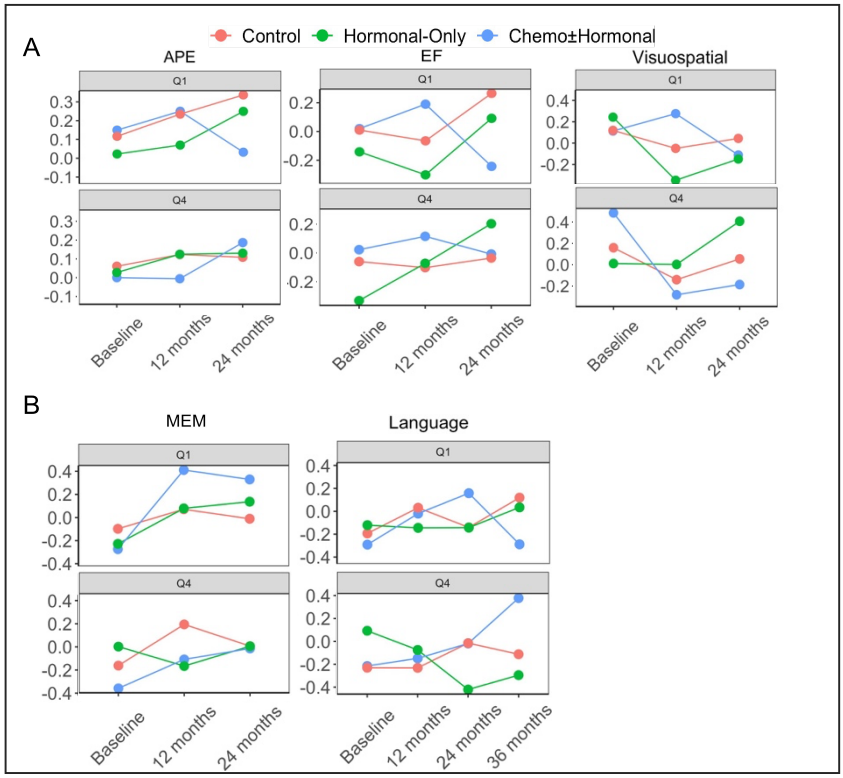

Figure 4.2: Adjusted mean cognitive scores over time for by PRS extreme quartiles Adjusted mean cognitive domain scores on the basis of least squares means from linear mixed-effects models show scores at baseline, 12 months, and 24 months for three treatment groups. **A)** Results based on $PRS_{(Immune\ Response)}$ for APE, EF, and visuospatial domain. **B)** Results based on $PRS_{(Endocytosis)}$ for Language and MEM.

compared to the *ε4-* survivors of the same group (P=3.7e-02). This observation replicates previous findings on *TLC* study, which observed decline in the score of *ε4+* hormonal-only

group in contrast to the corresponding *ε4-* group that showed an improvement in the same time window [105]. The aforementioned interactions remained insignificant on the 36-month data. No significant group-by-time-by-PRS interaction was observed for any of the pathway-PRSs or full-PRS neither on the 2-year nor on the 3-year data (Table 4.4).

4.3.3 EF

With exception the chemotherapy±hormonal group, the participants of the remaining groups exhibited the expected practice effects. Specifically, after a temporary, insignificant performance decline for all groups during the first year, the performance of the hormonal-only and the control groups exhibited significant improvements compared to the baseline (P<9.2e-03). Based on the 2-year follow up analysis, no significant group differences were observed longitudinally by *APOE* status (Table 4.4). However, at the third year, *ε4+* survivors in the hormonal-only group, had significantly lower EF levels compared to the *ε4-* survivors of the same group (P=3.8e-02). When accounting for the different PRS scores in the model, the 3-way interaction with the $PRS_{(Immune\ Response)}$ reached nominal significance in both 24-month and 36-month analyses (Table 4.4). A more detailed examination of the results was performed by including the $PRS_{(Immune\ Response)}$ quartile interaction with time and group (P=8.6e-03, based on the 24-month data). At the end of the two-year period, both the control and hormonal-only groups had improved performance compared to the baseline. Although the 2-year change for the low-risk hormonal-only group was similar to that of the control group (Figure 4.2) we observed a statistically significant result only for the latter group (0.25 units increase for the controls and 0.23 units increase for the hormonal-only group). In contrast to the other two groups, the

combined therapy participants of both PRS quartiles, exhibited a deterioration in their performance (Table 4.5, Figure 4.2). Finally, when we repeated the analysis on the 3-year data, were the interaction of $PRS_{(Immune\ Response)}$ quartile with time and group remained significant (Table 4.4).

Table 4.5: Marginal cognition changes by PRS-quartiles and treatment groups. The results are based on 2-year data.

				Change in the adjusted marginal mean (*P*-value) [12]		
Domain	Pathway	PRS quartile	Group	Baseline -12 months	12 months - 24 months	Baseline - 24 months
Immune response	APE	Q1	Controls	-0.12 (0.03)		-0.22 (4.3e-05)
			Hormonal-Only		-0.18 (0.03)	-0.22 (2.6e-03)
			Chemo±Hormonal			
		Q4	Controls			
			Hormonal-Only			
			Chemo±Hormonal			
	EF	Q1	Controls		-0.33 (3.5e-04)	-0.25 (6.9e-03)
			Hormonal-Only		-0.40 (2.9e-03)	
			Chemo±Hormonal			
		Q4	Controls			
			Hormonal-Only	-0.25 (0.05)		-0.52 (1.8e-04)
			Chemo±Hormonal			
	Visuo-spatial	Q1	Controls			
			Hormonal-Only	0.59 (1.7e-03)		
			Chemo±Hormonal			
		Q4	Controls			
			Hormonal-Only			
			Chemo±Hormonal	0.74 (9.2e-03)		
Endocytosis	Language	Q1	Controls	-0.21(4.8e-02)		
			Hormonal-Only			
			Chemo±Hormonal	-0.43(4.2e-02)		-0.67(3.8e-03)
		Q4	Controls			-0.22(4.3e-02)
			Hormonal-Only			0.51(8.1e-04)
			Chemo±Hormonal			
	MEM	Q1	Controls			
			Hormonal-Only	-0.31 (7.5e-03)		-0.34 (4.1e-03)
			Chemo±Hormonal	-0.68 (4.9e-04)		-0.60 (1.2e-02)
		Q4	Controls	-0.36 (1.1e-04)		
			Hormonal-Only			
			Chemo±Hormonal			

[1] Values represent the change in the scores between the two time points (Only significant changes are reported)
[2] *P*-values have been corrected for multiple comparisons using the Tukey formula
Green cells indicate improved scores; **Red** cells indicate decrease in the score

4.3.4 MEM

On average, all groups showed practice effects with marginal MEM scores at month 24 that outperformed the baseline levels. Interestingly, we observed a temporary decline during the first year in all three groups ($P<0.05$). The overall group-by-time interaction was insignificant, indicating similar rate of improvement among all participants (Table

4.4). Despite the insignificant three-way interaction *with APOE ε4* status, at month 12 we observed significantly lower value for the marginal MEM of the hormonal-only *ε4+* participants, compared to the *ε4-* participants of the same group (*P*=1.4e-02). Based on the 24-month dataset, we observed that $PRS_{(Endocytosis)}$ levels are significantly associated with the MEM score (*P*=3.7e-02, Table 4.4). The improvements observed at the end of the two-year period for all groups, were statistically significant for those with low $PRS_{(Endocytosis)}$ but not for the high-risk individuals (Table 4.5, Figure 4.2). In the 36-month dataset, the group-by-time-by-$PRS_{(Endocytosis)}$ interaction became insignificant.

4.3.5 VISUOSPATIAL

There was a deteriorating trajectory for all groups in the Visuospatial domain, especially for the hormonal-only group which had a significant reduction during the first year (*P*=1.7e-03). However, the overall rate of change of this domain showed no significant discrepancies among the groups (Table 4.4). Inclusion of *APOE* had no significant effect on that relationship either (Table 4.4). When we examined the effects of the pathway-PRSs we found a significant group-by-time-by-$PRS_{(Immune\ Response)}$ interaction (*P*=3.7e-02, Table 4.4). The effect of $PRS_{(Immune\ Response)}$ remained significant in the 36-month data (*P*=1.7e-02; Table 4.4). The visuospatial trajectory of the control and hormonal-only groups had a U shape exhibiting a temporary decline at the first year followed by an improvement at the second year (Figure 4.2, Table 4.5). The combined therapy group had an overall but insignificant decline in their scores, compared to the baseline (Table 4.5). Moreover, those in high-risk had a significant decline during the first year (Table 4.5). Any significant interaction between group, time, and the remaining of the pathway-PRSs did not survive

after removing *APOE* from the score (Table 4.4).

4.3.6 LANGUAGE

Overall, there was an increasing trajectory for all treatment groups without significant rate differences (Table 4.4). Pairwise comparisons revealed significantly lower levels for the hormonal-only and chemotherapy±hormonal groups, compared to the controls, at 24 and 36 months respectively (P=2.4e-02 for hormonal-only and P=3.6e-02 for combined therapy). *APOE* did not affect the rate of change between groups, as group-by-time-by-*APOE* was insignificant both in 24-month and 36-month data. After examining the effect of all seven pathway-PRSs and one overall-PRS in the score, we observed a marginal significance in the 36-month for the group-by-time-by-$PRS_{(Endocytosis)}$ interaction (Table 4.4, Figure 4.2). That was possible driven by the score difference between the extreme risk quartiles in the combined therapy group (Figure 4.2).

4.3 DISCUSSION

Our results indicated that, all treatment groups, with very few exceptions, exhibited the expected practice effects at the end of the 2-year follow up period, in most cognitive domains. Previous findings from the *TLC* study [105] that identified a decline for the combined therapy group in the APE domain were not replicated here. However, we did observe a decline for the combined group in the EF domain which is highly correlated with APE. Lower LM scores after hormonal therapy initiation for *APOE* ε4 carriers has also been observed previously from the same study [105] and this result was replicated in our work. In general, carrying at least one *APOE ε4* allele had a temporary, negative effect on

the performance of a limited number of domains, for those prescribed hormonal-only therapy. The main effect of *APOE ε4* was observed on the visuospatial domain but only when *APOE* was part of a pathway-PRS. The significance of this effect was lost after removing *APOE* from these scores. That confirms past findings that linked the presence of *APOE ε4* with lower visual memory and spatial ability among cancer survivors [113].

Incorporating pathway-PRSs in the analysis, expanded on our prior study by investigating dementia-related genetic markers beyond *APOE* [105]. Accounting for the aggregated effect of these variants in the form of pathway-PRSs helped revealing changes in EF, APE and visuospatial domains, suggesting inferior performance for survivors that are prescribed chemotherapy compared to hormonal therapy. The immune-response and endocytosis pathway-PRSs effect on specific cognitive domains, are meaningful in the sense that, in cancer there have been observations of disruption of these two pathways [111, 130, 131]. Existing literature supports the observed relation between the $PRS_{(Immune\ Response)}$ with executive function and visuospatial ability as several of the genes that compose the $PRS_{(Immune\ Response)}$ have been found to have a significant association with these domains. Specifically, CR1, EPHA1 and a number of MEF2C variants have been linked to visuospatial ability [132-135], whereas CLU to executive function [136, 137]. Supporting literature about the observed relation between language and memory with $PRS_{(Endocytosis)}$ has been published in the past, reporting significant links between $PRS_{(Endocytosis)}$ genes with recall and memory. Some of these findings include the association of CD2AP with delayed recall [138], PICALM and BIN1 with episodic memory [139-141], and several SORL1 variants with spatial abilities and episodic memory [142].

Several important limitations should be noted. First, as this is the first report associating PRS with cognitive outcomes in older women with BC, an independent replication sample was not available. Despite this limitation, the PRS employed here were derived from an independent large Alzheimer's and aging study. Second, statistical power was limited by available TLC cohort sample size. This problem is exacerbated when participants are subdivided into genetic risk quartiles, especially for the chemotherapy group which was the smallest group. Third, the follow-up of 3 years may not have captured longer term changes in cognitive functioning. Third, our results may not be generalizable to other populations for two reasons: first, the TLC participants are well educated, and cognitive reserve has been linked to post-treatment in the past [106] and second, to avoid the confounding effect of population stratification, only WNH individuals with available genotype were analyzed. Fourth, our results might be biased due to the impact of the practice effect on the cognitive scores. Previous study has shown that failing to adjust for the practice effect in a model can result to associations that are misleading and even inverse the direction of the relation [143]. Fifth, additional biological insight could have been gained, and stronger results may have been observed by incorporating a larger number of AD-related genes in the PRSs, based on findings from more recent and much larger GWASs.

In summary, genetic risk in relevant biological pathways was associated with post-treatment performance of older BC survivors on specific cognitive domains. A negative pathway-related genetic risk on cognitive outcome might be more pronounced for those

women treated with systemic chemotherapy. Pathway-level PRS may enhance our understanding on the biological process underlying cognitive changes in different domains and support treatment decisions. Further analysis is required to understand what is the common biological link between the cognitive domains that were associated to each of these pathway-PRSs. A possible explanation would be that immune-response PRS represents the risk for increased microglia activation in brain regions that are responsible for the executive function and visuospatial ability [144-149]. Similarly, language and memory changes might have a link to endocytosis induced synapse decline in the regions responsible for these two domains [150-153]. Overall, dementia-related genetic risk beyond *APOE* may be a useful tool in the clinicians' hands for assessing the likelihood of post-treatment cognitive deficiencies and assist decisions about the treatment type and duration. It can as well assist the investigation of the biological link between dementia-related pathways and cancer-indued cognitive changes.

Chapter 5

CONCLUSIONS

5.1 DISCUSSION

Since 1906 when the first case of Alzheimer's disease was observed, important domain progress has been made and valuable knowledge has been gained. Rapid technological advancements have led to significant genetic discoveries that shed light to many aspects of the disease. However, our understanding about the exact biological mechanism that leads to AD onset, is still limited. Appraisal of previously published literature revealed an explosion in the number of publications focusing on case-control PRS in AD, and an underutilization of biologically targeted PRSs. Currently, case-control PRS is mainly treated as mean for risk assessment and biomarker prediction. The usefulness of these applications is indisputable, but they have limited potential to promote knowledge about disease pathogenesis. Whereas in the near past, lack of data resources would possibly prevent the use of PRS towards that direction, current data availability provides exciting opportunities to unfold the potential of PRS in understanding the process of disease's pathogenesis and progression. These data resources should be utilized to develop hypothesis-driven PRSs. This is especially important in order to progress towards precision medicine solutions. In that context, exploration and understanding of the disease risk, biomarker heterogeneity, and progress should be approached by utilizing polygenic scores designated to answer the specific research questions. The genetic information integrated in a PRS, and its interpretation is highly linked to the GWAS on which the score was based on. Thus, risk-based PRS should not be treated as panacea but rather restrict its use to the corresponding task. In the present study, observation of significant associations between

endophenotype-PRSs and multiple disease outcomes, including AD risk and biomarker trajectories, provide an affirmation of the predictive ability of these scores. An interesting point resulting by these findings, is that the genetic complexity of the PRS should be analogous to the number of pathways that are involved to each outcome. The stronger the link of an outcome with specific pathways, the less complex PRS is required. Any additional SNPs probably introduce information from irrelevant to the outcome pathways, and thus decrease the prediction accuracy of the score. This is an indication that improved clinical utility may be achievable by controlling the genetic information that enters the PRS. The observed association between CSF amyloid and tau biomarkers with the corresponding endophenotype-PRSs, rather than the combined-PRS, is indicative of the potential of these scores to capture the genetic complexity of a response. Consequently, the targeted genetic information in the endophenotype-PRSs can mitigate the efforts of understanding the pathways related to that response. Further supporting evidence regarding the enhanced interpretability and performance of the biological-relevant PRS, emanates from the observed significant associations between specific cognitive domains with pathway-PRSs, but not with the overall-PRS. This finding does not only confirm the previous inference regarding the relevance of the genetic information in the PRS and the performance of the score, but also designates biological mechanisms that could be related to each of these cognitive domains. Depending on a person's genetic risk on specific dementia-related pathways, cancer therapies may impact different domains of cognitive abilities. While seemingly, endophenotype and pathway PRSs are similar, they are serving different purposes. The approach the latter are developed constitutes them as great tools for gaining biological insights for a disease such as the implication of specific biological

pathways in disease risk. Despite their use for assessment of several dementia endpoints, pathway-PRSs are still risk-based scores and their utilization for studying relations beyond disease-risk might be compromised. Whilst pathway-PRSs can be very useful in revealing evidence of the underlying biological disease mechanism, biomarker-based PRSs can provide information regarding the progression of hallmark disease biomarkers. That is important not only for disentangling the highly variable disease profiles but could also provide critical patient information during the recruitment stage of clinical trial that focus on target specific biomarker. Whereas there is still work to be done before PRS can be efficiently used for treatment development, the road towards precision medicine has opened. By now, it should be clear that in addition to the case/control GWAS, biomarker and endophenotype GWAS are necessary for development of biologically relevant PRSs. Although, in dementia research, risk based GWASs have exhibited significant increase in their sample size, AD-specific biomarker GWAS are limited in number and in terms of sample size. Since the sample size of a GWAS is very critical as it determines the power of the polygenic score, there is a necessity for increasing the availability and the size of these studies. Beyond sample size this work highlighted several additional factors that can affect the accuracy and power of the score. Usually the existing methods have managed to deal with some of these factors but not all of them, leaving space for further methodological refinement of the PRS formulas. In recent years, several advances on this domain have been achieved such as development of sophisticated Bayesian-based PRS methods for handling LD structure including PRS-CS, SBayesR, MegaPRS and more. In some cases, the tradeoff of some of these methods could be the increased computational requirements, the need of additional data for parameter tunning, as well as the introduction of restricting distributional assumptions. Because

polygenic methods are not only response specific but also disease specific, the criteria for their selection and application should be based in the context of the disease of interest. Despite the progress in PRS models, their predictive ability is still limited and thus not applicable to clinical practice. In an effort to capture the missing heritability, machine learning approaches have been lately integrated in the development of PRSs. By not making any distributional assumptions and by being able to handle multi-dimensional data and data interactions machine learning models could be a promising alternative to the existing PRSs. These approaches, however, can be not only very demanding in terms of computational time and memory, but they might also raise another difficulty as well. That of the score's limited interpretability. This is an essential drawback that needs to take into account, as development of tailored treatment solutions is based on the understanding of the disease's mechanisms. The importance of PRS's interpretability is highlighted by the continuous efforts for its improvement through the development of methodologies that incorporate functional annotation and biological relevant information in the scores. It is obvious that no approach is free of limitations and thus one needs to decide which of these aspects matter more in addressing their scientific question. Another issue that has been raised over the years is the restriction of the genetic research mainly on White, non-Hispanic populations. As PRS is ancestry specific, it is indented to be used on target samples that have the same ancestry as the discovery sample. In order to alleviate this barrier, efforts for generating multiethnic PRSs have become more intense lately, however this should be only a temporary solution as development of diverse study cohorts is required.

5.2 LIMITATIONS

First, the endophenotype-PRS findings may not be generalizable to populations other than White non-Hispanic. To avoid confounding effect due to population stratification, any WNH participants have been excluded from the analysis. The lack of studies with a large AD biomarker collection limited this work in using a subset of ADNI for running endophenotype GWASs. Harmonization between multiple studies could support this effort. Developing endophenotype-PRSs based on larger GWASs will improve the generalizability of the scores and the resulting associations. Third, endophenotype-PRS findings need to be validated on a totally independent sample, as currently the replication was performed on a subset of ADNI. Fourth, development of vascular-PRS needs to be repeated based on a study that provides information on multiple cardiovascular biomarkers. ADNI, is not ideal for that purpose as people with cardiovascular problems are excluded from the study. Fifth, performance comparison to the traditional risk-based PRSs is needed to further validate their effectiveness. The third aim also has several limitations that need to be addressed. First, the power to detect three-way interaction Time-Group-PRS was limited, especially for those who underwent chemotherapy. Second, the pathway-PRSs were based on the enrichment results of the 20 most significant genetic markers for AD. Expanding the number of SNPs in the pathway-PRSs, may lead to new findings and strengthen the already observed associations. Lastly, this was the first work to study the association between dementia-related pathway-PRS and post-treatment cognitive changes among women with BC and the study results need to be replicated in an independent cohort of women with BC.

5.3 FUTURE DIRECTIONS

This dissertation provides initial evidence of the potential of biological-relevant polygenic risk scores in Alzheimer's research. Novel endophenotype-PRSs were presented for which I provided evidence of their role as promising AD-specific markers. It was further shown that AD-specific genetic scores can expand the current research on cancer-related cognitive impairment, that has been so far studied only in relation to *APOE*.

The study of Alzheimer-specific pathway-PRSs in relation to cancer-cognition could also be further improved. Better predictive performance could be achieved by substituting the risk-based pathway-PRS weights by endophenotype-derived weights. Lastly, similar to the concept of biomarker-derived PRSs, dementia linked genetic information could be utilized to develop cognition- PRSs. It would be interesting to compare their accuracy levels with these of the pathway-PRSs, endophenotype- PRSs and risk-based PRSs.

Further refinement of the proposed endophenotype genetic scores can provide additional evidence of their effectiveness and generate new hypotheses. Because endophenotypes are closer to the biomarkers than risk is, genetic scores based on endophenotypes rather than risk-based polygenic scores could be more powerful tools for discovering new disease biomarkers. They may also assist with the recruitment of patients in clinical trials when the recruitment criteria are based on specific endophenotype risk that is considered to reflect the disease status. Furthermore, endophenotype-PRSs could be used to identify disease biomarker profiles and thus help understanding the heterogeneity of the disease. On individual level, classification of patients based on disease profiles could assist with

treatment decisions. In addition, as soon as therapies targeting particular pathways become available, they could be used as preventative tools by identifying individuals at increased genetic risk for AD pathology before the onset of clinical symptoms. Besides, their effectiveness in promoting the understanding on the Alzheimer's underlying mechanism should be tested. Specifically, pathway enrichment of endophenotype-PRS SNPs, could possibly provide insights on the disease's biological process. Finally, research efforts utilizing the recently developed concept of cell-specific-PRSs, could be complimented by the incorporation of endophenotype-PRSs. By investigating how the risk of different endophenotypes relates to the risk associated with different cell types (cell-specific-PRS), could enhance the knowledge regarding the disease pathogenesis. As an exemplar, disease risk or biomarker trajectory profiles derived based on endophenotype-PRSs could be studied in relation to the risk attributed to disease relevant cells, and possibly suggest candidate mechanisms of the disease pathogenesis.

5.4 SUMMARY

Complex diseases like Alzheimer's disease and other dementias are characterized by a complicated genetic component, increased phenotypic variability, and poorly understood pathogenesis mechanism. Polygenic risk scores are combinatorial measures expressing an individual's liability for a disease and they are perceived as promising tools in the precision medicine endeavor. According to the current practice they are developed based on case-control genome wide association studies and are most frequently used for risk prediction and prognosis. However, a rigorous literature review revealed that case-control PRSs have low interpretability and limited potential to provide insights in the disease's underlying

mechanism. This work investigated the potential of polygenic scores that integrate biologically relevant information to leverage disease knowledge beyond risk assessment and thus, support personalized solutions. Two types of PRSs were employed in this work, the endophenotype-PRS and the pathway-PRS. Both PRSs have a clear biological interpretation as their SNPs are either strongly linked to a homogeneous set of biomarkers (endophenotype-PRS) or to a dementia pathway (pathway-PRS). When responses of interest have a close relation to a biological function then, biologically informed PRS are preferred compared to the overall PRS. Data from the ANDI study indicated that endophenotype PRSs are preferred for prediction of CSF amyloid and tau biomarkers, whereas a combined PRS is preferred for neurodegeneration biomarkers, cognition, and disease risk. Implementation of dementia-related pathway-PRSs on the *TLC* study indicated that cancer-induced cognitive performance in specific domains is linked to immune and endocytosis related genetic risk, but not to the overall genetic risk. Both findings imply superiority of biologically targeted PRSs compared to overall PRS confirming previous studies that suggest that additional SNPs can harm the performance of the score. They also indicate that the "best" PRS is highly dependent to the research question. This work provides some initial, encouraging results about the role of biologically targeted PRSs in future research. Endophenotype and pathway PRSs can provide valuable information on the biology of the disease and thus, support the efforts for drug development. Also paired with case-control PRSs, which can provide an initial risk evaluation, endophenotype-PRSs could be used to further brake down disease profile and assist with treatment allocation.

REFERENCES

[1] Craig J. Complex diseases: Research and applications. Nature Education 2008;1:184.

[2] 2021 Alzheimer's disease facts and figures. Alzheimers Dement. 2021;17:327-406.

[3] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020;396:413-46.

[4] Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101:5-22.

[5] Chasioti D, Yan J, Nho K, Saykin AJ. Progress in Polygenic Composite Scores in Alzheimer's and Other Complex Diseases. Trends Genet. 2019;35:371-82.

[6] Fisher RA. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh. 1918;52:399-433.

[7] Craig J. Complex diseases: research and applications. Nature Education. 2008;1:184.

[8] Price AL, Spencer CC, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. Proc Biol Sci. 2015;282:20151684.

[9] Carter CO. Genetics of common disorders. Br Med Bull. 1969;25:52-7.

[10] Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. J Natl Cancer Inst. 2017;109.

[11] Li J, Holm J, Bergh J, Eriksson M, Darabi H, Lindstrom LS, et al. Breast cancer genetic risk profile is differentially associated with interval and screen-detected breast

cancers. Ann Oncol. 2015;26:517-22.

[12] Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, et al. Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. Ann Intern Med. 2009;150:541-50.

[13] International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460:748-52.

[14] Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, et al. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. Biol Psychiatry. 2017;81:470-7.

[15] Desikan RS, Fan CC, Wang Y, Schork AJ, Cabral HJ, Cupples LA, et al. Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. PLoS Med. 2017;14:e1002258.

[16] Mormino EC, Sperling RA, Holmes AJ, Buckner RL, De Jager PL, Smoller JW, et al. Polygenic risk of Alzheimer disease is associated with early- and late-life processes. Neurology. 2016;87:481-8.

[17] Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013;9:e1003348.

[18] Karlson EW, Chibnik LB, Kraft P, Cui J, Keenan BT, Ding B, et al. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Ann Rheum Dis. 2010;69:1077-85.

[19] Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med. 2016;375:2349-

58.

[20] Nalls MA, Escott-Price V, Williams NM, Lubbe S, Keller MF, Morris HR, et al. Genetic risk and age in Parkinson's disease: Continuum not stratum. Mov Disord. 2015;30:850-4.

[21] Broce IJ, Tan CH, Fan CC, Jansen I, Savage JE, Witoelar A, et al. Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. Acta Neuropathol. 2018.

[22] Dima D, Breen G. Polygenic risk scores in imaging genetics: Usefulness and applications. J Psychopharmacol. 2015;29:867-71.

[23] Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet. 2016;17:392-406.

[24] Mistry S, Harrison JR, Smith DJ, Escott-Price V, Zammit S. The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: Systematic review. Schizophr Res. 2017.

[25] Santoro ML, Moretti PN, Pellegrino R, Gadelha A, Abilio VC, Hayashi MA, et al. A current snapshot of common genomic variants contribution in psychiatric disorders. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2016;171:997-1005.

[26] 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;55:1068-87.

[27] Belsky DW, Sears MR, Hancox RJ, Harrington H, Houts R, Moffitt TE, et al. Polygenic risk and the development and course of asthma: an analysis of data from a four-

decade longitudinal study. The Lancet Respiratory Medicine. 2013;1:453-61.

[28] Layton J, Li X, Shen C, de Groot M, Lange L, Correa A, et al. Type 2 Diabetes Genetic Risk Scores Are Associated With Increased Type 2 Diabetes Risk Among African Americans by Cardiometabolic Status. Clin Med Insights Endocrinol Diabetes. 2018;11:1179551417748942.

[29] Ridge PG, Mukherjee S, Crane PK, Kauwe JS, Alzheimer's Disease Genetics C. Alzheimer's disease: analyzing the missing heritability. PLoS One. 2013;8:e79771.

[30] De Jager PL, Chibnik LB, Cui J, Reischl J, Lehr S, Simon KC, et al. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. Lancet Neurol. 2009;8:1111-9.

[31] Reeves GK, Travis RC, Green J, Bull D, Tipper S, Baker K, et al. Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. JAMA. 2010;304:426-34.

[32] Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet. 2010;376:1393-400.

[33] Raynor LA, Pankow JS, Rasmussen-Torvik LJ, Tang W, Prizment A, Couper DJ. Pleiotropy and pathway analyses of genetic variants associated with both type 2 diabetes and prostate cancer. Int J Mol Epidemiol Genet. 2013;4:49-60.

[34] Rodriguez-Rodriguez E, Sanchez-Juan P, Vazquez-Higuera JL, Mateo I, Pozueta A, Berciano J, et al. Genetic risk score predicting accelerated progression from mild cognitive impairment to Alzheimer's disease. J Neural Transm (Vienna). 2013;120:807-12.

[35] Harris SE, Davies G, Luciano M, Payton A, Fox HC, Haggarty P, et al. Polygenic risk

for Alzheimer's disease is not associated with cognitive ability or cognitive aging in non-demented older people. J Alzheimers Dis. 2014;39:565-74.

[36] Adams HH, de Bruijn RF, Hofman A, Uitterlinden AG, van Duijn CM, Vernooij MW, et al. Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. Alzheimers Dement. 2015;11:1277-85.

[37] Martiskainen H, Helisalmi S, Viswanathan J, Kurki M, Hall A, Herukka SK, et al. Effects of Alzheimer's disease-associated risk loci on cerebrospinal fluid biomarkers and disease progression: a polygenic risk score approach. J Alzheimers Dis. 2015;43:565-73.

[38] Vivot A, Glymour MM, Tzourio C, Amouyel P, Chene G, Dufouil C. Association of Alzheimer's related genotypes with cognitive decline in multiple domains: results from the Three-City Dijon study. Mol Psychiatry. 2015;20:1173-8.

[39] Yin X, Cheng H, Lin Y, Wineinger NE, Zhou F, Sheng Y, et al. A weighted polygenic risk score using 14 known susceptibility variants to estimate risk and age onset of psoriasis in Han Chinese. PLoS One. 2015;10:e0125369.

[40] Holm J, Li J, Darabi H, Eklund M, Eriksson M, Humphreys K, et al. Associations of Breast Cancer Risk Prediction Tools With Tumor Characteristics and Metastasis. J Clin Oncol. 2016;34:251-8.

[41] Pihlstrom L, Morset KR, Grimstad E, Vitelli V, Toft M. A cumulative genetic risk score predicts progression in Parkinson's disease. Mov Disord. 2016;31:487-90.

[42] Gibson J, Russ TC, Adams MJ, Clarke TK, Howard DM, Hall LS, et al. Assessing the presence of shared genetic architecture between Alzheimer's disease and major depressive disorder using genome-wide association data. Transl Psychiatry. 2017;7:e1094.

[43] Lacour A, Espinosa A, Louwersheimer E, Heilmann S, Hernandez I, Wolfsgruber S,

et al. Genome-wide significant risk factors for Alzheimer's disease: role in progression to dementia due to Alzheimer's disease among subjects with mild cognitive impairment. Mol Psychiatry. 2017;22:153-60.

[44] Lall K, Magi R, Morris A, Metspalu A, Fischer K. Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. Genet Med. 2017;19:322-9.

[45] Li H, Feng B, Miron A, Chen X, Beesley J, Bimeh E, et al. Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. Genet Med. 2017;19:30-5.

[46] Natarajan P, Young R, Stitziel NO, Padmanabhan S, Baber U, Mehran R, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. Circulation. 2017;135:2091-101.

[47] Oh JJ, Park S, Lee SE, Hong SK, Lee S, Kim TJ, et al. Genetic risk score to predict biochemical recurrence after radical prostatectomy in prostate cancer: prospective cohort study. Oncotarget. 2017;8:75979-88.

[48] Sengupta SM, MacDonald K, Fathalli F, Yim A, Lepage M, Iyer S, et al. Polygenic Risk Score associated with specific symptom dimensions in first-episode psychosis. Schizophr Res. 2017;184:116-21.

[49] Tan CH, Hyman BT, Tan JJX, Hess CP, Dillon WP, Schellenberg GD, et al. Polygenic hazard scores in preclinical Alzheimer disease. Ann Neurol. 2017;82:484-8.

[50] Ten Broeke SW, Elsayed FA, Pagan L, Olderode-Berends MJW, Garcia EG, Gille HJP, et al. SNP association study in PMS2-associated Lynch syndrome. Fam Cancer. 2017.

[51] Chaudhury S, Patel T, Barber IS, Guetta-Baranes T, Brookes KJ, Chappell S, et al.

Polygenic risk score in postmortem diagnosed sporadic early-onset Alzheimer's disease. Neurobiol Aging. 2018;62:244 e1- e8.

[52] Hindy G, Wiberg F, Almgren P, Mealander O, Melander M. Polygenic Risk Score for Coronary Heart Disease Modifies the Elevated Risk by Cigarette Smoking for Disease Incidence. Circ Genom Precis Med. 2018;11.

[53] Logue MW, Panizzon MS, Elman JA, Gillespie NA, Hatton SN, Gustavson DE, et al. Use of an Alzheimer's disease polygenic risk score to identify mild cognitive impairment in adults in their 50s. Mol Psychiatry. 2018.

[54] Paul KC, Schulz J, Bronstein JM, Lill CM, Ritz BR. Association of Polygenic Risk Score With Cognitive Decline and Motor Progression in Parkinson Disease. JAMA Neurol. 2018.

[55] Seibert TM, Fan CC, Wang Y, Zuber V, Karunamuni R, Parsons JK, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. BMJ. 2018;360:j5757.

[56] Tan CH, Fan CC, Mormino EC, Sugrue LP, Broce IJ, Hess CP, et al. Polygenic hazard score: an enrichment marker for Alzheimer's associated amyloid and tau deposition. Acta Neuropathol. 2018;135:85-93.

[57] Sabuncu MR, Buckner RL, Smoller JW, Lee PH, Fischl B, Sperling RA, et al. The association between a polygenic Alzheimer score and cortical thickness in clinically normal subjects. Cerebral cortex. 2012;22:2653-61.

[58] Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31:1466-8.

[59] Mak TS, Kwan JS, Campbell DD, Sham PC. Local True Discovery Rate Weighted

Polygenic Scores Using GWAS Summary Data. Behav Genet. 2016;46:573-82.

[60] So HC, Sham PC. Improving polygenic risk prediction from summary statistics by an empirical Bayes approach. Sci Rep. 2017;7:41262.

[61] Chouraki V, Reitz C, Maury F, Bis JC, Bellenguez C, Yu L, et al. Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. J Alzheimers Dis. 2016;53:921-32.

[62] Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am J Hum Genet. 2015;97:576-92.

[63] Hu Y, Lu Q, Powles R, Yao X, Yang C, Fang F, et al. Leveraging functional annotations in genetic risk prediction for human complex diseases. PLoS computational biology. 2017;13:e1005589.

[64] Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. 1996;58:22.

[65] Golan D, Rosset S. Effective genetic-risk prediction using mixed models. Am J Hum Genet. 2014;95:383-93.

[66] Abraham GK, A.; Zobel, J.; Inouye, M.; . Performance and Robustness of Penalized and Unpenalized Methods for Genetic Prediction of Complex Human Disease. Genetic epidemiology. 2013;37:184–95.

[67] de Vlaming R, Groenen PJ. The Current and Future Use of Ridge Regression for Prediction in Quantitative Genetics. Biomed Res Int. 2015;2015:143712.

[68] Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genetic epidemiology. 2017;41:469-80.

[69] Zhou X, Carbonetto P, Stephens M. Polygenic Modeling with Bayesian Sparse Linear Mixed Models. PLoS Genet. 2013;9.

[70] Rakitsch B, Lippert C, Stegle O, Borgwardt K. A Lasso multi-marker mixed model for association mapping with population structure correction. Bioinformatics. 2013;29:206-14.

[71] Sleegers K, Bettens K, De Roeck A, Van Cauwenberghe C, Cuyvers E, Verheijen J, et al. A 22-single nucleotide polymorphism Alzheimer's disease risk score correlates with family history, onset age, and cerebrospinal fluid Abeta42. Alzheimers Dement. 2015;11:1452-60.

[72] So HC, Sham PC. Exploring the predictive power of polygenic scores derived from genome-wide association studies: a study of 10 complex traits. Bioinformatics. 2017;33:886-92.

[73] Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet. 2013;45:400-5, 5e1-3.

[74] Darst BF, Koscik RL, Racine AM, Oh JM, Krause RA, Carlsson CM, et al. Pathway-Specific Polygenic Risk Scores as Predictors of Amyloid-beta Deposition and Cognitive Function in a Sample at Increased Risk for Alzheimer's Disease. J Alzheimers Dis. 2017;55:473-84.

[75] Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. Mol Psychiatry. 2017.

[76] Tosto G, Bird TD, Tsuang D, Bennett DA, Boeve BF, Cruchaga C, et al. Polygenic risk scores in familial Alzheimer disease. Neurology. 2017;88:1180-6.

[77] Cruchaga C, Del-Aguila JL, Saef B, Black K, Fernandez MV, Budde J, et al. Polygenic risk score of sporadic late-onset Alzheimer's disease reveals a shared architecture with the familial and early-onset forms. Alzheimers Dement. 2018;14:205-14.

[78] Escott-Price V, Sims R, Bannister C, Harold D, Vronskaya M, Majounie E, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. Brain. 2015;138:3673-84.

[79] Escott-Price; V. M, A.J; Huentelman, M.; Hardy, J. Polygenic Risk Score Analysis of Pathologically Confirmed Alzheimer Disease. Annals of Neurology. 2017;82:311-4.

[80] Lupton MK, Strike L, Hansell NK, Wen W, Mather KA, Armstrong NJ, et al. The effect of increased genetic risk for Alzheimer's disease on hippocampal and amygdala volume. Neurobiol Aging. 2016;40:68-77.

[81] Harrison TM, Mahmood Z, Lau EP, Karacozoff AM, Burggren AC, Small GW, et al. An Alzheimer's Disease Genetic Risk Score Predicts Longitudinal Thinning of Hippocampal Complex Subregions in Healthy Older Adults. eNeuro. 2016;3.

[82] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45:1452-8.

[83] Lee SH, Harold D, Nyholt DR, Consortium AN, International Endogene C, Genetic, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. Hum Mol Genet. 2013;22:832-41.

[84] Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease. Transl Psychiatry. 2018;8:99.

[85] Fisher R. The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh. 1919;2:399-433.

[86] Shen L, Thompson PM. Brain Imaging Genomics: Integrated Analysis and Machine Learning. Proc IEEE Inst Electr Electron Eng. 2020;108:125-62.

[87] Daunt P, Ballard CG, Creese B, Davidson G, Hardy J, Oshota O, et al. Polygenic Risk Scoring is an Effective Approach to Predict Those Individuals Most Likely to Decline Cognitively Due to Alzheimer's Disease. J Prev Alzheimers Dis. 2021;8:78-83.

[88] Banks SJ, Qiu Y, Fan CC, Dale AM, Zou J, Askew B, et al. Enriching the design of Alzheimer's disease clinical trials: Application of the polygenic hazard score and composite outcome measures. Alzheimers Dement (N Y). 2020;6:e12071.

[89] Escott-Price V, Myers A, Huentelman M, Shoai M, Hardy J. Polygenic Risk Score Analysis of Alzheimer's Disease in Cases without APOE4 or APOE2 Alleles. J Prev Alzheimers Dis. 2019;6:16-9.

[90] Chaudhury S, Brookes KJ, Patel T, Fallows A, Guetta-Baranes T, Turton JC, et al. Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. Transl Psychiatry. 2019;9:154.

[91] genoSCORE-LAB.

[92] Ahmad S, Bannister C, van der Lee SJ, Vojinovic D, Adams HHH, Ramirez A, et al. Disentangling the biological pathways involved in early features of Alzheimer's disease in the Rotterdam Study. Alzheimers Dement. 2018;14:848-57.

[93] Desikan RS, Schork AJ, Wang Y, Thompson WK, Dehghan A, Ridker PM, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. Circulation. 2015;131:2061-9.

[94] Sierksma A, Escott-Price V, De Strooper B. Translating genetic risk of Alzheimer's disease into mechanistic insight and drug targets. Science. 2020;370:61-6.

[95] Alzheimer's Disease Neuroimaging Initiative.

[96] Xu Y, Goodacre R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. J Anal Test. 2018;2:249-62.

[97] Jack CR, Jr., Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14:535-62.

[98] Jack CR, Jr., Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology. 2016;87:539-47.

[99] Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society. 1996;58:267-88.

[100] Zhao Y, Dantony E, Roy P. Optimism Bias Correction in Omics Studies with Big Data: Assessment of Penalized Methods on Simulated Data. OMICS. 2019;23:207-13.

[101] Harrison JR, Mistry S, Muskett N, Escott-Price V. From Polygenic Scores to Precision Medicine in Alzheimer's Disease: A Systematic Review. J Alzheimers Dis. 2020;74:1271-83.

[102] Barber RC, Phillips NR, Tilson JL, Huebinger RM, Shewale SJ, Koenig JL, et al. Can Genetic Analysis of Putative Blood Alzheimer's Disease Biomarkers Lead to Identification of Susceptibility Loci? PLoS One. 2015;10:e0142360.

[103] Ahles TA, Root JC, Ryan EL. Cancer- and cancer treatment-associated cognitive change: an update on the state of the science. J Clin Oncol. 2012;30:3675-86.

[104] van der Willik KD, Schagen SB, Ikram MA. Cancer and dementia: Two sides of the same coin? Eur J Clin Invest. 2018;48:e13019.

[105] Mandelblatt JS, Small BJ, Luta G, Hurria A, Jim H, McDonald BC, et al. Cancer-Related Cognitive Outcomes Among Older Breast Cancer Survivors in the Thinking and Living With Cancer Study. J Clin Oncol. 2018:JCO1800140.

[106] Ahles TA, Saykin AJ, McDonald BC, Li Y, Furstenberg CT, Hanscom BS, et al. Longitudinal assessment of cognitive changes associated with adjuvant treatment for breast cancer: impact of age and cognitive reserve. J Clin Oncol. 2010;28:4434-40.

[107] Plun-Favreau H, Lewis PA, Hardy J, Martins LM, Wood NW. Cancer and neurodegeneration: between the devil and the deep blue sea. PLoS Genet. 2010;6:e1001257.

[108] Roe CM, Fitzpatrick AL, Xiong C, Sieh W, Kuller L, Miller JP, et al. Cancer linked to Alzheimer disease but not vascular dementia. Neurology. 2010;74:106-12.

[109] Behrens MI, Lendon C, Roe CM. A common biological mechanism in cancer and Alzheimer's disease? Curr Alzheimer Res. 2009;6:196-204.

[110] Roe CM, Behrens MI, Xiong C, Miller JP, Morris JC. Alzheimer disease and cancer. Neurology. 2005;64:895-8.

[111] Nudelman KNH, McDonald BC, Lahiri DK, Saykin AJ. Biological Hallmarks of Cancer in Alzheimer's Disease. Mol Neurobiol. 2019;56:7173-87.

[112] Harrison RA, Rao V, Kesler SR. The association of genetic polymorphisms with neuroconnectivity in breast cancer patients. Sci Rep. 2021;11:6169.

[113] Ahles TA, Saykin AJ, Noll WW, Furstenberg CT, Guerin S, Cole B, et al. The relationship of APOE genotype to neuropsychological performance in long-term cancer survivors treated with standard dose chemotherapy. Psychooncology. 2003;12:612-9.

[114] Lehrer S, Rheinstein PH. Breast Cancer, Alzheimer's Disease, and APOE4 Allele in the UK Biobank Cohort. J Alzheimers Dis Rep. 2021;5:49-53.

[115] Smith CJ, Ashford JW, Perfetti TA. Putative Survival Advantages in Young Apolipoprotein varepsilon4 Carriers are Associated with Increased Neural Stress. J Alzheimers Dis. 2019;68:885-923.

[116] Lo MT, Kauppi K, Fan CC, Sanyal N, Reas ET, Sundar VS, et al. Identification of genetic heterogeneity of Alzheimer's disease across age. Neurobiol Aging. 2019;84:243 e1- e9.

[117] Bonham LW, Geier EG, Fan CC, Leong JK, Besser L, Kukull WA, et al. Age-dependent effects of APOE epsilon4 in preclinical Alzheimer's disease. Ann Clin Transl Neurol. 2016;3:668-77.

[118] Tesi N, van der Lee SJ, Hulsman M, Jansen IE, Stringa N, van Schoor NM, et al. Immune response and endocytosis pathways are associated with the resilience against Alzheimer's disease. Transl Psychiatry. 2020;10:332.

[119] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

[120] Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet. 2005;76:887-93.

[121] Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284-7.

[122] McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279-83.

[123] Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O, et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. Mol Psychiatry. 2011;16:903-7.

[124] Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet. 2009;41:1094-9.

[125] Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA. 2010;303:1832-40.

[126] Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 2011;43:429-35.

[127] Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet. 2011;43:436-41.

[128] International Genomics of Alzheimer's Disease C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. Alzheimers Dement. 2015;11:658-71.

[129] Guerreiro R, Bras J, Hardy J. SnapShot: genetics of Alzheimer's disease. Cell. 2013;155:968- e1.

[130] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646-74.

[131] Ahles TA, Saykin AJ. Candidate mechanisms for chemotherapy-induced cognitive

changes. Nat Rev Cancer. 2007;7:192-201.

[132] Kamboh MI, Fan KH, Yan Q, Beer JC, Snitz BE, Wang X, et al. Population-based genome-wide association study of cognitive decline in older adults free of dementia: identification of a novel locus for the attention domain. Neurobiol Aging. 2019;84:239 e15-e24.

[133] Sunderaraman P, Cosentino S, Schupf N, Manly J, Gu Y, Barral S. MEF2C Common Genetic Variation Is Associated With Different Aspects of Cognition in Non-Hispanic White and Caribbean Hispanic Non-demented Older Adults. Front Genet. 2021;12:642327.

[134] Chibnik LB, Shulman JM, Leurgans SE, Schneider JA, Wilson RS, Tran D, et al. CR1 is associated with amyloid plaque burden and age-related cognitive decline. Ann Neurol. 2011;69:560-9.

[135] Keenan BT, Shulman JM, Chibnik LB, Raj T, Tran D, Sabuncu MR, et al. A coding variant in CR1 interacts with APOE-epsilon4 to influence cognitive decline. Hum Mol Genet. 2012;21:2377-88.

[136] McFall GP, Sapkota S, McDermott KL, Dixon RA. Risk-reducing Apolipoprotein E and Clusterin genotypes protect against the consequences of poor vascular health on executive function performance and change in nondemented older adults. Neurobiol Aging. 2016;42:91-100.

[137] Sapkota S, McFall GP, Masellis M, Dixon RA. A Multimodal Risk Network Predicts Executive Function Trajectories in Non-demented Aging. Front Aging Neurosci. 2021;13:621023.

[138] Smith JA, Zhao W, Yu M, Rumfelt KE, Moorjani P, Ganna A, et al. Association Between Episodic Memory and Genetic Risk Factors for Alzheimer's Disease in South

Asians from the Longitudinal Aging Study in India-Diagnostic Assessment of Dementia (LASI-DAD). J Am Geriatr Soc. 2020;68 Suppl 3:S45-S53.

[139] Liu Z, Dai X, Zhang J, Li X, Chen Y, Ma C, et al. The Interactive Effects of Age and PICALM rs541458 Polymorphism on Cognitive Performance, Brain Structure, and Function in Non-demented Elderly. Mol Neurobiol. 2018;55:1271-83.

[140] Barral S, Bird T, Goate A, Farlow MR, Diaz-Arrastia R, Bennett DA, et al. Genotype patterns at PICALM, CR1, BIN1, CLU, and APOE genes are associated with episodic memory. Neurology. 2012;78:1464-71.

[141] Greenbaum L, Ravona-Springer R, Lubitz I, Schmeidler J, Cooper I, Sano M, et al. Potential contribution of the Alzheimer's disease risk locus BIN1 to episodic memory performance in cognitively normal Type 2 diabetes elderly. Eur Neuropsychopharmacol. 2016;26:787-95.

[142] Reynolds CA, Zavala C, Gatz M, Vie L, Johansson B, Malmberg B, et al. Sortilin receptor 1 predicts longitudinal cognitive change. Neurobiol Aging. 2013;34:1710 e11-8.

[143] Cerulla N, Arcusa A, Navarro JB, de la Osa N, Garolera M, Enero C, et al. Cognitive impairment following chemotherapy for breast cancer: The impact of practice effect on results. J Clin Exp Neuropsychol. 2019;41:290-9.

[144] Passamonti L, Rodriguez PV, Hong YT, Allinson KSJ, Bevan-Jones WR, Williamson D, et al. [(11)C]PK11195 binding in Alzheimer disease and progressive supranuclear palsy. Neurology. 2018;90:e1989-e96.

[145] Wang XL, Li L. Microglia Regulate Neuronal Circuits in Homeostatic and High-Fat Diet-Induced Inflammatory Conditions. Front Cell Neurosci. 2021;15:722028.

[146] Levit A, Regis AM, Gibson A, Hough OH, Maheshwari S, Agca Y, et al. Impaired

behavioural flexibility related to white matter microgliosis in the TgAPP21 rat model of Alzheimer disease. Brain Behav Immun. 2019;80:25-34.

[147] Suridjan I, Pollock BG, Verhoeff NP, Voineskos AN, Chow T, Rusjan PM, et al. In-vivo imaging of grey and white matter neuroinflammation in Alzheimer's disease: a positron emission tomography study with a novel radioligand, [18F]-FEPPA. Mol Psychiatry. 2015;20:1579-87.

[148] Henkel K, Karitzky J, Schmid M, Mader I, Glatting G, Unger JW, et al. Imaging of activated microglia with PET and [11C]PK 11195 in corticobasal degeneration. Mov Disord. 2004;19:817-21.

[149] Gibson EM, Monje M. Microglia in Cancer Therapy-Related Cognitive Impairment. Trends Neurosci. 2021;44:441-51.

[150] Burrinha T, Martinsson I, Gomes R, Terrasso AP, Gouras GK, Almeida CG. Upregulation of APP endocytosis by neuronal aging drives amyloid-dependent synapse loss. J Cell Sci. 2021;134.

[151] Morrison JH, Baxter MG. The ageing cortical synapse: hallmarks and implications for cognitive decline. Nat Rev Neurosci. 2012;13:240-50.

[152] Nguyen LD, Ehrlich BE. Cellular mechanisms and treatments for chemobrain: insight from aging and neurodegenerative diseases. EMBO Mol Med. 2020;12:e12075.

[153] Forrest MP, Parnell E, Penzes P. Dendritic structural plasticity and neuropsychiatric disease. Nat Rev Neurosci. 2018;19:215-34.

[154] Yeh HH, Ogawa K, Balatoni J, Mukhapadhyay U, Pal A, Gonzalez-Lepera C, et al. Molecular imaging of active mutant L858R EGF receptor (EGFR) kinase-expressing nonsmall cell lung carcinomas using PET/CT. Proc Natl Acad Sci U S A. 2011;108:1603-

8.

CURRICULUM VITAE

Danai Chasioti

Education

2022	PhD, Bioinformatics	Indiana University-Purdue University Indianapolis, USA/School of Informatics & Computing
2015	MS, Biostatistics	Indiana University-Purdue University Indianapolis, USA/School of Public Health
2011	MS, Biostatistics	National & Kapodistrian University of Athens, Greece/School of Medicine
2007	BS, Statistics & Actuarial Science	University of the Aegean, Greece/ School of Science

Experience

- Scientist I, Human Genetics — Biogen Inc. — 2021-current

- Research assistant — Dr. Andrew J. Saykin's Lab — 2018-2021
  - Development of biomarker-based Polygenic Risk Scores for Alzheimer's disease and other dementias

- Research assistant — Dr. Li Shen's Lab — 2015-2018
  - Study high-order, directional drug interaction adverse effects using data-mining techniques on electronic health records
  - Imaging genetics in AD

- Biostatistics Consultant — Community Health Network — 2014-2015
  - Identify factors that affect the Case Mix Index of the Community Health Network

- Research assistant — Dr. Constantin T. Yiannoutsos — 2011-2014
  - Identify the factors that affect the Case Mix Index of hospitals in the U.S., using longitudinal analysis

- Teaching assistant — 2011-2012
  Department of Mathematics, Indiana University-Purdue University Indianapolis, USA

- Statistics Tutor — 2011
  National & Kapodistrian University of Athens, Greece

- Laboratory assistant, Biostatistics I 2009
  Harokopion University of Athens, Greece

Publications and Presentations (chronologically)

[1] Chasioti, D., Jacobson, T., Nho K., Risacher, S.L., Gao, S., Yan, J., and Saykin, A.J. (2021). Biomarker-based polygenic risk scores for profiling genetic susceptibility in Alzheimer's disease. Alzheimer's Association Int'l Conf. 2021.

[2] Chasioti, D., Jacobson, T., Nho, K., Risacher, S.L., Gao, S., Yan, J., and Saykin, A.J. (2020). Endophenotype driven polygenic risk scores for Alzheimer's disease. Alzheimer's Dement., 16:e046766. https://doi.org/10.1002/alz.046766

[3] Chasioti, D., Yan, J., Nho, K., and Saykin, A.J. (2019). Progress in Polygenic Composite Scores in Alzheimer's and Other Complex Diseases. Trends in genetics, 35(5):371–382.

[4] Chasioti, D., Yao, X., Zhang, P., Lerner, S., Quinney, S. K., Ning, X., Li, L., and Shen, L. (2019). Mining Directional Drug Interaction Effects on Myopathy Using the FAERS Database. IEEE Journal of Biomedical and Health Informatics, 23(5):2156–2163

[5] Chasioti, D., Yan, J., Nho, K. and Saykin, A.J. (2019). Polygenic composite scores in Alzheimer's disease: a systematic review. Alzheimer's and Dementia, 15:631-632. https://doi.org/10.1016/j.jalz.2019.06.2554

[6] Chasioti, D., Jacobson, T., Nho, K., Risacher, S.L., Yan, J. and Saykin, A.J. (2019). Biomarker-based polygenic risk scores for Alzheimer's disease. Alzheimer's and Dementia, 15:284-285. https://doi.org/10.1016/j.jalz.2019.06.683

[7] Liu, K., Yao, X., Yan, J., Chasioti, D., Risacher, S.L., Nho, K., Saykin, A.J, and Shen, L. (2017). Alzheimer's Disease Neuroimaging Initiative. Transcriptome-Guided Imaging

Genetic Analysis via a Novel Sparse CCA Algorithm. Graphs Biomed. Image Anal. Comput. Anal. Imaging Genet., 10551:220-229. doi:10.1007/978-3-319-67675-3_20

[8] Chasioti, D., Yao, X., Zhang, P., Ning, X., Li, L., and Shen, L. (2017). Mining directional drug interaction effects on myopathy using the FAERS database. PSB'17: Pac. Symp. Biocomp., Hawaii, USA. 9. Chasioti, D., Yao, X., Zhang, P., Quinney, S.K., Ning, X., Li, L., and Shen, L. (2017)

[9] Mining and visualizing the network of directional drug interaction eects. NetSci'17: Int'l School and Conf. on Network Science, Indianapolis, USA.

[10] Chasioti, D., Yao, X., Zhang, P., Lerner, S., Quinney, S.K., Ning, X., Li, L., and Shen, L. (2017). Mining directional drug interaction effects on myopathy using the FAERS database. 7th CTSI Symp. on Disease and Therapeutic Response Modeling and Simulation, Indianapolis, USA