Pathak, Gita A. **Genetic characterization of comorbidity patterns in aging associated diseases using integrative genomics**. Doctor of Philosophy (Biomedical Sciences)
August 2019, 231 pages, 2 tables, 37 figures, 229 references, 52407 words

ABSTRACT

The aging population in the US continues to grow at an exponential rate estimated to reach more than 90 million by 2060. The coexistence of two or more diseases (comorbidity) is prevalent in ages 65 years and above, and the number of comorbidities increases with age. The genetic factors underlying presence and absence of comorbidities is a severely understudied research domain. Alzheimer's disease (AD) is a type of dementia affecting 5.5 million people with an average age of diagnosis at 70 years. Hypertension is a coexisting condition in 60% AD individuals, also known as direct comorbidity. On the other hand, cancer is reported to be inversely comorbid with AD; individuals with cancer history have been reported to have lower risk of AD and vice versa. Furthermore, individuals with cancer history are diagnosed with long term side effects of radiation therapy – radiotoxicity. Twin-based studies have reported that certain gene variants are associated with radiotoxicity phenotypes with a heritability of 66%.

This study proposes to investigate genetic factors associated with the direct and inverse comorbidity of AD with hypertension and cancer, and proctitis – a radiotoxicity phenotype observed in survivors of prostate cancer. The study aims to integrate gene variants, derived-gene expression and copy number variation (CNV), followed by functional and pathway-based prioritization of observed findings. We used genome-wide and cerebral spinal fluid profile to investigate presence of hypertension with AD to evaluate individual-level differences, followed by targeted investigation of neighboring gene expression profiles of identified variants. We found several novel genes associated with AD-hypertension comorbidity. The investigation between AD and cancer identified regions in chromosomes 4, 5 and 19 that are targeted by miRNA-17 family along with other miRNAs reported to be inversely expressed and play opposite role in pathogenicity of both diseases. The SNP-derived transcriptomic profile between AD and cancer highlighted involvement of sirtuin signaling. The findings together indicate involvement of mitochondrial and metabolic dysregulation which possibly contribute in differences of the epithelial-mesenchymal-transition. The SNP-derived expression and CNV association with proctitis highlighted genes involved in DNA-repair and mitochondrial ROS damage pathways.

KEYWORDS

Alzheimer's Disease, breast, prostate, cancer, hypertension, miRNA, proctitis

GENETIC CHARACTERIZATION OF COMORBIDITY PATTERNS IN AGING ASSOCIATED DISEASES USING INTEGRATIVE GENOMICS

Gita A. Pathak, M.Sc.

APPROVED:

_____
Major Professor

_____
Committee Member

_____
Committee Member

_____
Committee Member

_____
University Member

_____
Graduate Advisor

_____
Chair, Department of Microbiology, Immunology & Genetics

_____
Dean, Graduate School of Biomedical Sciences

GENETIC CHARACTERIZATION OF COMORBIDITY PATTERNS IN AGING ASSOCIATED DISEASES USING INTEGRATIVE GENOMICS

DISSERTATION

Presented to the Graduate Council of the

University of North Texas Health Science Center at Fort Worth in

Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

By:

Gita A. Pathak, M.Sc.

August 6, 2019

ACKNOWLEDGMENTS

When I first applied to UNTHSC, I didn't know what a truly enriching and learning experience it would become. Today, as I am completing my journey, I would like to express my gratitude to all the inspiring people who augmented my personal and academic growth.

I am deeply appreciative to my advisor, Dr. Nicole Phillips. She is the epitome of a true mentor, her encouragement, patience, support, guidance and training has truly transformed my professional aspirations. She took me under her wing in January 2018, in the middle of my program. She never asked me to drop my original project even though it seemed as a 'dead end'. On the contrary, she gave me her brain child – Alzheimer's vs cancer, as a side-project to give me motivation. Little did I know that it would turn into a fascinating project. Dr. Phillips is an inspiration to several students who aspire to have research careers. While I regret that I couldn't spend more time under her tutelage, I am proud to be called her student for life. I am grateful for the teachings I received from my former PhD mentor – (late) Dr. Ranajit Chakraborty, his genius and humility will forever be my inspiration. I am so grateful to him for accepting me, I am honored to be called his student.

Next, I want to thank my committee members, Drs. Robert Barber, John Planz, Zhengyang Zhou, and Ignacy Gryczynski. Dr. Barber has been always been a source of support, encouraging students to ask questions and maintaining an open-door policy for students in need. I am deeply appreciative that he supported my academic journey with constructive and valuable feedback. He also provided with collaborative projects which helped accrue my skills. I am grateful to him for providing me with funding support for the last year of the PhD program. I am very grateful to Dr.Planz for being my graduate advisor and committee member. Dr.Planz interviewed me for the PhD program, and without his confidence in me I wouldn't have reached this academic pedestal in my life. He also provided me with multiple teaching opportunities, for which I am truly grateful. Dr. Zhou joined my committee last year, and his contribution and support has resulted in tremendous improvement in my projects. He was always available to answer my questions and gave positive feedback for my projects. Dr. Zhou's constant encouragement has played huge role in the success of my PhD dissertation. In the end, I want to express my gratitude to Dr. Gryczynski, he is the best university member any student can ever ask for. His undeterred support and encouragement have been instrumental in the success of my graduate journey. He always checked in with my progress and made sure that I was keeping up with my timeline to graduate on time. Without the support of my committee members, I would not have made this far; thank you for believing in me and my ability to finish my PhD project.

I want to thank my family for their perennial support and encouragement to complete my PhD degree. My mother, and my grandparents always envisioned me to be a scholar. While I am deeply saddened that they are not here with me today, their passing has given me purpose and the passion to serve others. My research interests are and will always be in their honor. I want to thank my aunt, Anju Rattan who called me every day to check on me, and often wondered, how have I not

finished my PhD already. She's more excited than I am to see me complete this journey. I want to thank my uncle, Dr. Ashish Dwivedi who always believed in me, and gave me the advice and support to become independent. Lastly, I want to thank my brothers – Vikram and Arjun Rattan, they have tolerated me for many years, but they deserve my sincerest gratitude for tolerating me the last few years when I became a little more insane than usual. My big brother, Arjun gave me the extra push I needed to reach where I am today. He has seen me grow both professionally and personally. He never shies from reminding me that I should be appreciative of overcoming my own struggles, for they have made me a better person.

I want to thank my best friend, Amreen Farooqui, friends - Jeremy Julian, Shruti Patil, Kirthikaa Balapattabi, Frank Wendt, Navita Lopez, Talisa Silzer, Viviana Mancilla, Danielle Reid, Casandra Setser, Bing Song, and so many more. I have leaned on their shoulders for celebrations and crying about failures.

I would like to acknowledge my second family on campus – the career center team. I want to thank James Renfro, Nancy Eanes, Eronia King, and Lydia Negron, for their confidence in me has played a huge role in my success. I faced multiple breakdowns and backlogs during the start my journey. But, working for them kept me sane while I worked through my disappointments. I have built a cherished relationship that will last a lifetime with each one of them.

I want to thank Jie Sun, Carla Lee Johnson, Jacklyn Crisp, Deborah Turman, Deanna Ranker, Jessica Medina and all the staff members for their constant support.

In the end, I would like to acknowledge the funding support I received from GSBS Office, Dr. Meharvan Singh's Neurobiology of Aging NIH-T32 predoctoral fellowship for 2017-2018, and Dr. Robert Barber's Texas Alzheimer's Research and Care Consortium by the Darrell K Royal Texas Alzheimer's Initiative, directed by the Texas Council on Alzheimer's Disease and Related Disorders, and the HABLE consortium.

- Gita

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

List of Publications

In Revisions

Two-stage Bayesian GWAS of 9,638 individuals identifies SNP-regions that are targeted by miRNAs inversely expressed in Alzheimer's and cancer

Gita Pathak et. al – Journal of Alzheimer's & Dementia

Genome-wide methylation of mild cognitive impairment in Mexican Americans highlights genes involved in synaptic transport, AD-precursor phenotypes, and metabolic morbidities

Gita Pathak et. al – Journal of Alzheimer's Disease

Submitted

SNP-derived transcriptomics & copy number variation in proctitis— a radiotherapy side effect—points to altered mitochondrial and DNA repair mechanisms

Gita Pathak et. al – Scientific Reports

To be submitted

SNP-derived transcriptomics of multiple tissues between Alzheimer's and cancer implicates miR-17 family and sirtuin signaling

Gita Pathak et. al – Journal of Human Genetics

Integrative analysis of gene variants, transcriptome and methylation to understand Alzheimer's - hypertension comorbidity

Gita Pathak et. al – Genome Research

Patient genetics shape chronic wound microbiome composition and influences healing

*Craig D. Tipton, Randall D. Wolcott, Nicholas E. Sanford, Clint Miller, **Gita Pathak**, Jie Sun, Derek Fleming, Kendra P. Rumbaug, Todd D. Little, Nicole Phillips, Caleb Phillips*

Published

Talisa Silzer, Robert Barber, Jie Sun, **Gita Pathak**, Leigh Johnson, Sid O'Bryant, Nicole Phillips "Circulating mitochondrial DNA: new indices of type 2 diabetes-related cognitive impairment in Mexican Americans", PLOS ONE, (2019).

Frank Wendt, **Gita Pathak**, Antti Sajantila, Ranajit Chakraborty, Bruce Budowle. Global genetic variation of selection opiate metabolism genes in self-reported healthy individuals. Pharmacogenomics J. (2017).

Chapter 1

INTRODUCTION TO INVESTIGATION OF COMORBIDITY PATTERNS OF AGING-RELATED DISEASES – CANCER AND RADIOTOXICITY, HYPERTENSION AND ALZHEIMER'S USING INTEGRATIVE GENOMICS

SUMMARY. The United States aging population is expected to double from 46 million in 2014 to more than 95 million by 2060. As aging itself is a risk factor for multiple chronic conditions, the incidence rates of age-related diseases, such as cancer, hypertension and Alzheimer's disease (AD), are the highest in the population ages 65 years and older. The percentage of individuals with two more chronic conditions (comorbidity) also increases with age in both men and women, with comorbidity rates of 62% in ages 65-74 years and up to 81.5% in ages 85 years and older. Comorbidity- defined as a medical condition that exists along with an index condition, is healthcare concern due to its high prevalence and our limited pathological understanding in the older population. While hypertension is known to co-occur with Alzheimer's, epidemiological studies have reported an inversely comorbid relationship between cancer and Alzheimer's disease. Inverse comorbidity is defined as lower-than-expected probability of a disease occurring in individuals who have been diagnosed with other medical conditions. This research study proposes to investigate the both direct and indirect patterns of comorbidity using genotype data from NCBI's Database of Genotype and Phenotype (dbGaP) and Alzheimer's disease and Neuroimaging Initiative (ADNI).

The strategy is to implement multivariate correlation and regression approaches, with the integration of reference transcriptomics data and multiple tissue-based network analysis. There are three integrated aims: 1) to identify genetic polymorphisms associated with cancer-related radiation treatment in prostate cancer survivors, 2) to investigate the impact of hypertension with Alzheimer's disease, and 3) to study the direction of effect of genetic variants in Alzheimer's and cancer that may be responsible for the inverse association. The clinical decision and management of elderly care is marred by issues of poor prognosis after treatment, increased cognitive decline with comorbid condition, and polypharmacy. As the amount of health data available is likely to grow exponentially, the integrative genomics approach will allow for a more fine-grained characterization of comorbidity spectra and will provide possibilities for understanding the molecular underpinning of disease co-occurrences, thereby contributing considerably to the landscape of precision treatment initiatives in the elderly.

1.1 SPECIFIC AIMS

There has been a significant shift in population demographic trends, with a projected doubling of the segment aged 65 and older in the U.S., from 46 million in 2014 to more than 95 million by 2060[1]. Aging itself is a risk factor to multiple chronic conditions, and the most prevalent age-associated diseases are Alzheimer's, cancer, and hypertension in the elderly population of United States[2]. By the year 2050, the number of Alzheimer's patients age 65 and older may nearly triple to 13.8 million, and cancer incidences for breast and prostate are expected to double[3,4]. With the increase in rate of cancer diagnosis, the number of cancer survivors is currently 12 million and will continue to rise in the United States[5]. Hypertension is (1) the most prevalent cardiovascular condition, affecting 250 million US individuals and its occurrence increases with age[6]; and (2) the most prevalent comorbid condition in Alzheimer's disease[7,8]. While Alzheimer's disease is directly comorbid with hypertension, several studies have reported an inverse comorbidity between cancer and Alzheimer's disease[9]. A multidimensional genetics approach has the potential to investigate comorbidity patterns due to confluences in transcriptomics, proteins, and genomics[10].

PROBLEM STATEMENT: In genetic association studies, the concept of genetic basis of comorbidity is often adjusted as covariate[11]; however, the genetic underpinnings remain understudied especially in the elderly population[12]. An in-depth investigation of genetic-variant influences on functional association[13] may shed much-needed light on the age-related pathophysiology of chronic conditions and their co-prevalence to better therapeutic approaches and clinical trial cohort design. **HYPOTHESIS:** Novel genetic risk factors underlie the observed patterns of co-occurrence in hypertension, cancer (and cancer therapy response), and Alzheimer's disease. **APPROACH:** To investigate these divergent comorbidity patterns, this study proposes to leverage the genetic data

available from public repositories combined with an integrative genomics approach to identify gene expression and pathway interaction among these age-associated diseases.

SPECIFIC AIM 1: To identify trends and association of genetic variants associated with adverse effects of radiation therapy in prostate cancer survivors. *Hypothesis:* Variants in genes responsible for DNA repair response and/or their associated pathways confer risk for radiosensitivity and increased severity of side effects. *Approach***:** Genetic data from prostate cancer individuals who have received radiation will be compared to age-matched controls with gene-based association study, followed by a multivariate GWAS to identify polymorphisms associated with the combination of sub-phenotypes of radiotoxicity.

SPECIFIC AIM 2: To investigate direct comorbidity of hypertension and Alzheimer's disease affected individuals. *Hypothesis*: Genetic variants associated with comorbid diseases confer higher risk by influencing changes at transcriptomic level. *Approach*: Using cross-disorder association analysis of genetic data from each phenotype, shared etiologies will be identified which are unique to individuals with comorbidity when compared to individuals diagnosed with only hypertension.

SPECIFIC AIM 3: To study genetic and functional networks to investigate possible inverse comorbidity pattern between Alzheimer's disease and cancer. *Hypothesis*: SNP markers have opposite effects with specific cancer-type and Alzheimer's, resulting in off-target effects in disease networks. *Approach*: Multinomial logistic regression will be used to identify genetic variants with opposite effects, followed by gene-based association and functional prioritization of observed variants.

1.2 BACKGROUND

The advancement of medicine better therapeutics and lifestyle intervention awareness have resulted in one of the most significant demographic shift observed in the United States. The growth of the population ages 65 and older, has been steadily increasing since the 1960's, and is projected to double from 46 million in 2014 to more than 98 million by 2060 (Figure 2)[1].

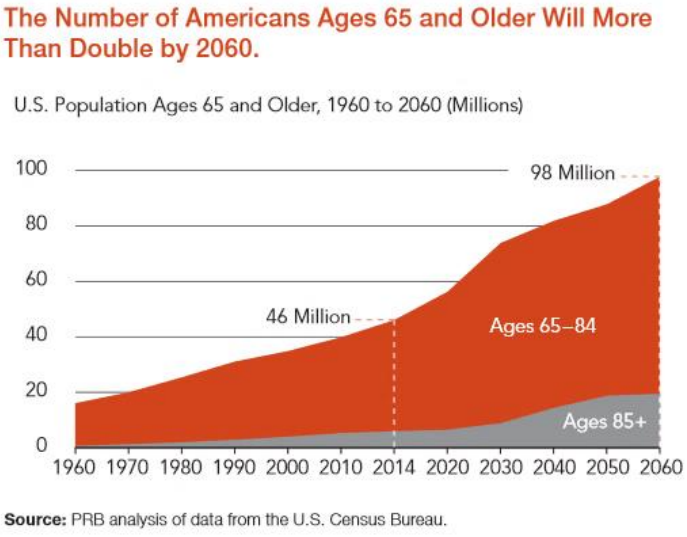


Figure 2: Incidence of aging population in the United States (Source: Population Reference Bureau, 2015)

This demographic shift places a global healthcare burden as aging is accompanied by increased incidences of chronic age-associated diseases, fueling the demand for more targeted medical interventions for the elderly[14].

1.2.1 Prevalent diseases in the aging population

The aging process is described as gradual decline in overall physiological functions and is known to present itself as a risk factor for many chronic diseases: cancer, metabolic disease and neurodegeneration[15]. Cancer refers to a group of diseases, wherein the body cell's which may originate in different tissues, start dividing rapidly without succumbing to self-regulated apoptosis[16]. Cancerous tumors may possess malignancy or benignity; malignant tumors spread into, or invade, nearby tissues and have been known to break off from the site of origin and travel to different parts of the body through the blood or lymph

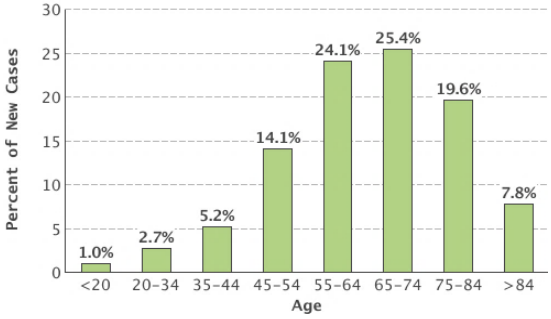


Figure 3: Percent of New Cancers by Age Group: All Cancer Sites (Source: SEER, 2007-2011)

nodes[17]. Mechanisms of aging and cancer development cross paths resulting in a time-dependent accumulation of cellular damage[18]. In 2018, an estimated number of cancer diagnosis are 1,735,350 and 609,640 people will die from the disease in the United States, and the worldwide cancer incidences are expected to rise to 21 million by 2030. Several cancers are age-dependent, whose median age at diagnosis varies from 61 years for breast cancer, 66 years for prostate cancer, 68 years for colorectal cancer, and 70 years for lung cancer. As the aging population continues to rise and with the high prevalence of cancer in the elderly population, a targeted investigation of cancer and associated morbidities in populations whose average age is 65 years and older is crucial[4].

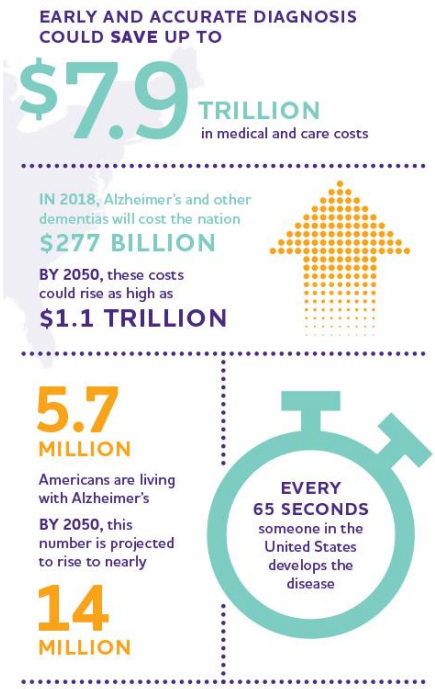


Figure 4: Incidence of Alzheimer's disease in the United States. (Source: Alzheimer's Association)

Another prevalent chronic condition in the aging population is Alzheimer's disease, affecting an estimated 5.5 million people aged 65 and older. Alzheimer's disease is a neurodegenerative condition resulting in an irreversible and gradual cognitive and functional decline, eventually restricting the ability to carry out routine tasks. Alzheimer's disease is the sixth-leading cause of death in the United States and the fifth-leading cause of death among those age 65 and older. It also is a chief cause of disability and poor health in older individuals[3]. There are multiple factors that been known to cause age-associated brain changes in Alzheimer's disease including genetics, family history, environmental, and lifestyle factors. Currently, there is no cure or treatment that limits the progression of cognitive decline in Alzheimer's affected individuals, putting a burden of $277 billion in medical and care costs, and is estimated to reach $1.1 trillion by 2050 (Figure 4)[19].

1.2.2 Comorbidity and its impact in the aging population

Aging is known to escalate the risk of developing concomitant chronic illnesses in older individuals creating a healthcare burden on the baby boomer generation[20]. Multimorbidity, the coexistence of two more chronic conditions has become widely prevalent, due to a decline in mortality rates and increase in the aging population (Figure 5)[7]. The concept of multimorbidity is a successor to comorbidity – defined in the 1970 as the presence of one or more diseases in addition to the primary disease in the same individual. Comorbidity and multimorbidity terms were originally derived to guide clinical diagnosis; treatment of comorbidity encompasses a primary diagnosis, while multimorbidity has a tailored approach to the patient's diagnosis in which no disease is defined as primary. Thus, the two concepts are not mutually exclusive but consider the co-presence of two chronic from different clinical perspectives. The terminologies – comorbidity and multimorbidity are often used interchangeably, but hereon we will be using the clinical definition of comorbidity, to signify the co-presence of a chronic disease with a primary condition[21]. An epidemiological study investigated the incidence rates of 15 different chronic conditions and the percentage of comorbidities in individuals aged 65 years and older enrolled as Medicare beneficiaries. More than two-thirds of the population had comorbidities, which increased with age to 62% in older individuals aged 65-74 years and 81.5% for those aged 85 years or more. The presence of comorbidity is over 90% in all prevalent chronic conditions fueling the need to investigate and

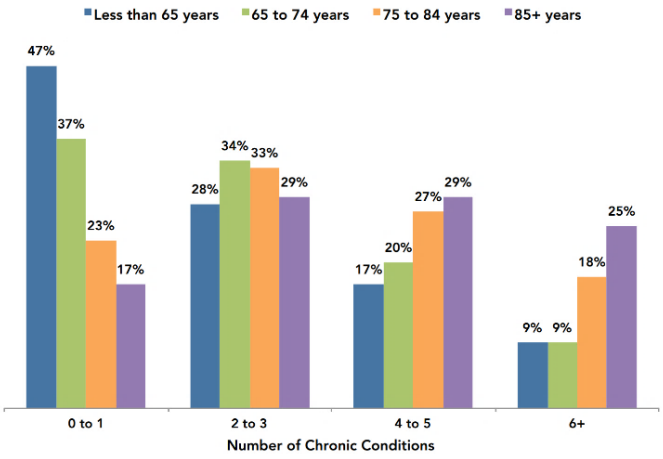


Figure 5: Percentage of Medicare FFS Beneficiaries by Number of Chronic Conditions and Age

(Source: Chronic Conditions Among Medicare Beneficiaries, Chartbook, 2012 edition)

understand the relationship between the co-presence of two or more diseases in the aging demographic[22].

With the rapid increase in prevalence of multiple chronic diseases in the aging population, comorbidity incidence rates in individuals with cancer aged ≥ 65 years is even higher when compared to an age-matched population without cancer. While rate of survival varies among cancer sub-types, there are currently more than 12 million cancer survivors aged ≥ 65 years (Table 1)[5] with little known understanding on how other factors such as comorbidities affect quality-of-life outcome after cancer treatment.

Over the last few years, the importance of comorbid conditions in clinical oncology[23] has gathered interest due to its impact on treatment associated toxicity in older adults[24]. More than half of the cancer types receive radiation therapy, often combined with chemotherapy and/or surgery, and with improvements in targeted delivery of radiation, radiotherapy outcomes have improved the overall survival rates. Even with the precision delivery of radiation, surrounding normal tissue may get irradiated and lead to acute or late side effects from radiotherapy[25]. It has been known that all individuals experience some level of toxicity, varying in severity (from minor to life-threatening) and in duration (from week to lifetime). There are many factors that may contribute to the

	MALE AND FEMALE			MALE			FEMALE		
	NO.	PERCENT	CUMULATIVE PERCENT	NO.	PERCENT	CUMULATIVE PERCENT	NO.	PERCENT	CUMULATIVE PERCENT
All Ages, y	15,533,220			7,377,100			8,156,120		
0–14	65,190	<1	<1	32,060	<1	<1	33,130	<1	<1
15–19	47,180	<1	1	23,610	<1	1	23,570	<1	1
20–29	187,490	1	2	90,730	1	2	96,760	1	2
30–39	408,790	3	5	166,170	2	4	242,620	3	5
40–49	958,600	6	11	347,700	5	9	610,900	7	12
50–59	2,389,670	15	26	963,410	13	22	1,426,260	17	30
60–69	4,141,950	27	53	2,027,150	27	49	2,114,800	26	56
70–79	4,011,790	26	79	2,148,940	29	79	1,862,850	23	79
≥80	3,322,560	21	100	1,577,330	21	100	1,745,230	21	100

Table 1: Estimated Number of US Cancer Survivors as of January 1, 2016, by Sex and Age at Prevalence. (Source: Miller et.al)

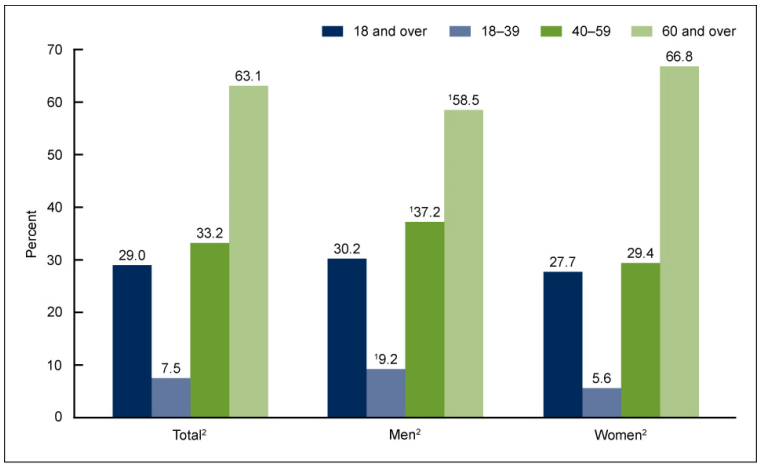
individual's susceptibility to radiation side-effects, including cancer type, tumor site, age, comorbidities and genetics[26]. The role of comorbidities in overall survival rates is consistent among various age-dependent cancers, such as prostate and breast cancer. In a study investigating comorbidity as predictor of overall survival in prostate cancer individuals treated with radiotherapy reported that comorbidity index was a stronger and independent predictor over age in a retrospective study spanning over 10 years[27]. The relationship between comorbidities and severity of radiation toxicity in older individuals remains ambiguous, due to the involvement of various factors such as severity of comorbidity, possible underlying genetic pathologies of complex diseases and lack of availability of research studies investigating the patients with and without comorbidities[28].

Similar to cancer, comorbidity burden in individuals affected with Alzheimer's is far greater than in age-matched individuals without Alzheimer's[29]. Increase in comorbidities among Alzheimer's population has been associated with increased levels of cognitive decline as observed on mini-mental state examination adjusting for age, gender, education and care setting[29]. Comorbidities have also been known to be associated with lower interest in personal care, decreased agility, thus affecting everyday performance and independence[30]. The incidence of comorbidities present in the Alzheimer's or other dementia-affected population is summarized in Table 2. Most of the comorbidities are vascular risk factors, with hypertension being most prevalent in the elderly individuals aged 65 and above with Alzheimer's[3].

Table 2: Percentages of Medicare beneficiaries aged ≥65 years with Alzheimer's disease (AD) and other dementias by specified coexisting medical conditions (Source: Alzheimer's Association Report 2011)

Coexisting condition	Percentage with AD or other dementia and the coexisting condition
Hypertension	60%
Coronary heart disease	26%
Stroke—late effects	25%
Diabetes	23%
Osteoporosis	18%
Congestive heart failure	16%
Chronic obstructive pulmonary disease	15%
Cancer	13%
Parkinson's disease	8%

1.2.3 Hypertension as a comorbidity in Alzheimer's disease



[1]Men significantly different from women in the same age group.
[2]Significant increasing trend by age.
NOTES: Estimates for age group 18 and over are age adjusted by the direct method to the 2000 U.S. Census population using age groups 18–39, 40–59, and 60 and over. Crude estimates for age group 18 and over are 32.1%, total; 31.8%, men; and 32.4%, women. Access data table for Figure 1 at: https://www.cdc.gov/nchs/data/databriefs/db289_table.pdf#1.
SOURCE: NCHS, National Health and Nutrition Examination Survey, 2015-2016.

Figure 6: Prevalence of hypertension among adults aged 18 and over, by sex and age: United States, 2015–2016 (Source: CDC)

Hypertension affects over 250 million people in the Americas, with an overall prevalence of 29%. This increases with age, with 63% of individuals aged 60 and above being affected (Figure 6)[6]. Hypertension is clinically defined as systolic blood pressure over 140mmHg or diastolic blood pressure greater than or equal to 90 mmHg or currently taking medication to lower high blood pressure[31]. Hypertension remains a major healthcare challenge because it is not only a risk factor for cardiovascular diseases but is also the most prevalent comorbid condition in the elderly population as summarized in and Table 2.

Hypertension is also the most prevalent comorbidity in Alzheimer's disease affecting 60% individuals (Table 2). Studies have proven that there is a strong association between hypertension as a risk factor for Alzheimer's disease[32]. Uncontrolled hypertension has been known to disrupt the blood-brain barrier around the brain capillaries resulting in dysfunction of subcortical vessels impeding the delivery of nutrients. Additionally, hypertension is implicated in endothelial damage to the brain via nitric oxide production, triggering inflammatory response and promoting plaque formation – one of the hallmarks of Alzheimer's disease[29]. Association studies in Alzheimer's subjects with hypertension observed poor performance on all cognition tests, supporting the idea that vascular dysfunctions affect cognitive decline [33]. Hypertension and Alzheimer's disease also share a genetic polymorphism of Apolipoprotein E4 allele, Sery O. et al reported a 1.5-fold

increased risk to Alzheimer's disease in subjects with hypertension and ApoE4 allele[34]. Furthermore, several studies have indicated that hypertension medications such as the angiotensin-receptor blockers have protective effects, concluded from MRI studies showing higher volume of hippocampal and parenchymal region and improved performance on cognition assessment tests[35]. These findings suggest a critical need to investigate the genetic underpinnings of hypertension as a comorbidity in age-related diseases.

1.2.4 Alzheimer's disease and cancer: inverse comorbidity

While Alzheimer's disease and cancer share a direct comorbidity association with hypertension, several studies have reported an inverse comorbidity between cancer and Alzheimer's disease[36]. Inverse comorbidity is defined as lower-than-expected probability of a disease occurring in individuals who have been diagnosed with other medical conditions[21]. More than a decade ago, Roe *et al.* found an inverse risk of cancer development in Alzheimer's participants, and lower risk of Alzheimer's development in participants with cancer history[37]. In their follow-up report in 2010, the study confirmed inverse association of sporadic Alzheimer's and cancer, but not with vascular dementia[38]. White *et al.* reported a similar finding of reduced risk of Alzheimer's in individuals with non-melanoma skin cancer in a longitudinal study with the average age 79 years of study participants[39].

However, some studies have reported conflicting findings for inverse association between cancer and Alzheimer's disease quoting survival bias as the reason for inverse association rather than a true inverse relationship. In an association study conducted in Medicare population of over 800,000 cancer cases, reported a moderate inverse relationship suspecting ascertainment bias or diagnostic misclassification as probable source of contradictory results[40]. Another study, also did not find association between early-stage cancer and risk of Alzheimer's, but supported the findings

between aggressive cancer diagnosis and inverse incidence of Alzheimer's disease[41]. Additionally, if the inverse relation between Alzheimer's and cancer is due to competing risk of death, then similar risk would be expected with other age-related diseases; however in a 2013 study conducted using SEER data found the only inverse relationship, between Alzheimer's and Parkinson's with cancer out of 400 disease pair associations[42]. Recently, in a retrospective study of more than 3 million US veterans-cancer survivors aged ≥65 years were found to be at a lower risk of Alzheimer's disease, than other age-related outcomes. The researchers also found the association of cancer treatment with lower risk of Alzheimer's disease. The odds ratio was 0.89 in 14 cancer types after excluding prostate, colorectal and melanoma cancer[43]. In a meta-analysis study of association studies from 1966 to 2013, Alzheimer's individuals had a decreased incident cancer by 42% and individuals with cancer history had 37% reduced risk of Alzheimer's disease[44]. These epidemiological findings have also been investigated using gene expression and microRNA data, and have observed overexpression of Pin, p53, and Wnt signalling pathway in cancer, but decreased expression in Alzheimer's, suggesting a possible mechanism behind the inverse relationship[45-48]. Additionally, functional analysis of differentially expressed genes revealed oxidative phosphorylation in the mitochondria to be inversely regulated in Alzheimer's and lung cancer, but down regulated in both glioblastoma and Alzheimer's disease leading to a state of chronic inflammation[49]. Epidemiological investigations into other neurodegenerative diseases – Parkinson's, have also reported an inverse association with cancer[50]. Altogether, these findings necessitate the identification of potentially divergent genetic processes to design individualized therapeutic strategies for Alzheimer's disease. Exploring genetic transformations during cancer and its treatment would also establish risk profiles for neurodegenerative diseases and identify

possible off-targets of drugs/underlying pathology providing protective effects against Alzheimer's disease and improving quality-of-life outcome after cancer treatments.

1.2.5 Studying comorbidity of complex diseases

Comorbidity research is a complex domain, and although aging is naturally inherent to comorbidity there are other factors that influence the wide spectrum of comorbidity characterization[51]. A mapping of two complex diseases in an individual's state of disease can be viewed in comorbidity space (Figure 7), where the axes represent quantitative measures influenced by environmental factors, genetic variants, stress and/or therapeutic interventions. The co-presence of two complex diseases suggest a shared pathology pathway, which can be environmental and/or genetic, or in the case of inverse comorbidity, may be due to an off-target effect from another disease state or treatment. To investigate these shared interactions, several approaches using health care data, to create networks of disease co-occurrence or using molecular procedures to identify functional products associated with disease phenotypes have been employed. These two-dimensional approaches clearly seem insufficient to understand the complexity of age-associated diseases since as networks of diseases change depending on variant, its protein-product, and its role in different tissues, giving rise to multiple models of multifunctionality[52].

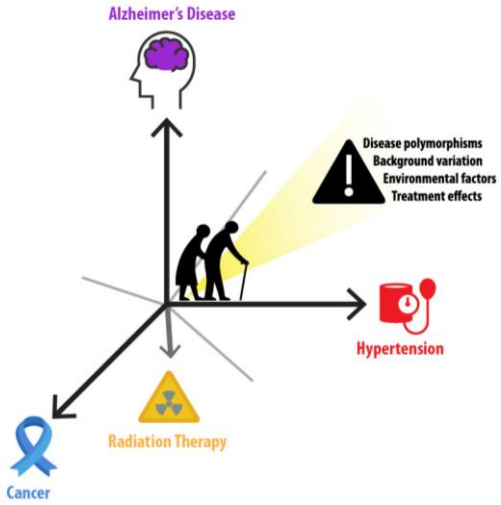


Figure 7: Comorbidity space of an individual's state of disease.
The axes represent disease outcomes and lighter bars between the axes are possible genetic polymorphisms and other factors that may be shared between two disease outcomes

A multidimensional genetics approach has the potential to investigate comorbidity patterns, as gene variants and gene products affect the stationary and dynamic spectrum of disease states[53].

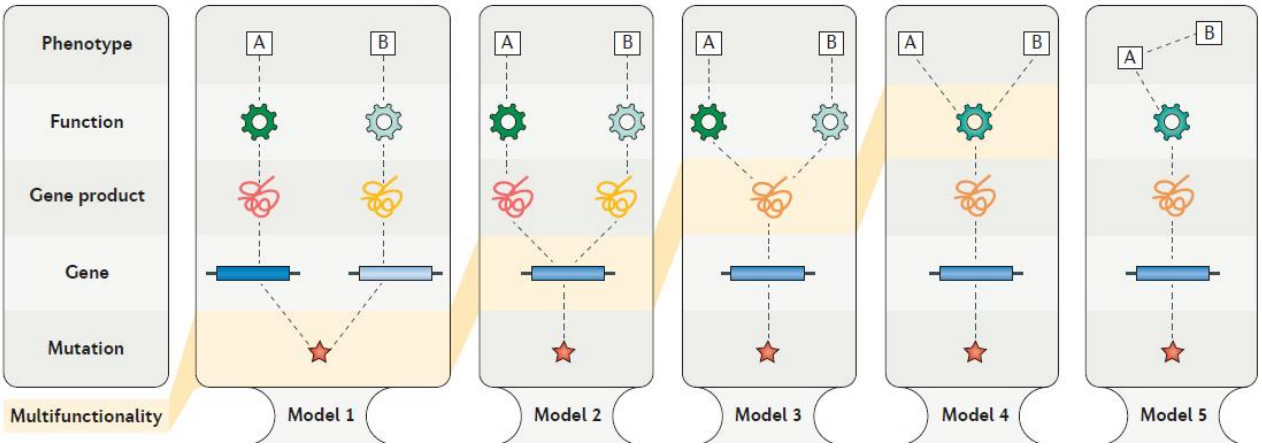


Figure 8: Models of multifunctionality and pleiotropy. (Source: Xin Hu et al.)
Model 1: A genetic variant affects expression of more than one gene; Model 2: gene variant affects gene expression with multiple functional products; Model 3: The gene product has multiple functions, possibly in different tissues; Model 4: The same gene variant, expression and protein function involved in two phenotypes; Model 5: Mediated Pleiotropy- The effect on phenotype-B is mediated through phenotype-A, but the genetic variants or product are not associated with phenotype-B.

Disease complexity and comorbid states can be understood in the light of genetic variants of disease under pleiotropic and multifunctionality models. Summarized in Figure 8, these models elucidate the influence of genetic variants as polymorphisms that affect multiple genes and/or gene expression of more than one gene. Protein-products of genes contain domain of different functions or may have functional differences based on tissue types. A protein product may also exert direct influence over two phenotypes. The goal of an integrative genomics approach is to 1) embrace the development and progression of complex diseases and their co-presence due to confluences in transcriptomics, proteins, and genomics; and 2) to comprehensively catalog the entire spectrum of genomic and proteomic interaction networks in different complex disease states[10,21].

1.3 SIGNIFICANCE

With the increase in the aging population coupled with high incidence rates of chronic conditions and comorbidities, investigation of comorbidity is a critical public healthcare demand[54]. The development of comorbidities as an adverse effect of cancer treatment limits the quality of life for cancer survivors. In patients of Alzheimer's, presence of comorbidity- hypertension - has been independently associated with cognitive decline[29]. The role of pleiotropy in co-presence of complex diseases is pertinent to understand the relationship between different chronic conditions and associated risk profiles[21].

INNOVATION: To date, genome wide association studies (GWAS) have successfully implicated several genetic variants in Mendelian disorders and many candidate loci for complex diseases. However, due to the multifactorial nature of complex diseases owing to polygenic and environmental influence, comparing allele frequencies between cases and controls, as done in GWAS, provides limited information[55]. Hypertension is a common comorbidity to Alzheimer's[34], but cancer and Alzheimer's are inversely comorbid[56]. This research study aims to investigate multiple genetic dimensions of cancer, cancer-related radiation treatment, hypertension and Alzheimer's disease in the aging population. Using this integrative genomics approach, computational tools and publicly available genetic data of disease-affected individuals will be analyzed to study multiple aspects of variation[53]– from SNPs, to protein expression levels, to pathway states– that are often altered in diseases. Leveraging multi-omics to understand network biology behind complex diseases in the context of comorbidity will allow for transformation of the large genomic data into biological insight that improves our capacity to better diagnose, treat and hopefully prevent complex diseases[57].

CONTRIBUTION: Comorbidities affect clinical decision and management and patients of advanced age with comorbidities are less likely to receive aggressive treatment for their cancer than individuals without comorbid conditions[12]. The results from this study will contribute to the literature of genetic variants, expression levels and pathway interaction in the comorbid conditions in the elderly population. The integrative genomics including functional analysis of genetic data based on gene-sets, eQTL expression levels, in multiple tissue types, and network analysis in individual-disease and comorbidity-association studies will potentially explain the crossover influence between different chronic conditions[58,59]. This research study not only establishes a pipeline for analysis of comorbidity in complex disease, but it aims to specifically address the degree of functional overlap and molecular relationship among three of the most prevalent chronic conditions (hypertension, Alzheimer's disease, and cancer/cancer treatments), to inform better screening profiles, diagnostic tests and therapeutics strategies in the understudied aging population.

1.4 PROBLEM STATEMENT & HYPOTHESIS

PROBLEM STATEMENT: In genetic association studies, the concept of genetic basis of comorbidity is often adjusted as covariate; however, the genetic underpinnings remains understudied in the elderly population. Due to the multifactorial etiology of complex disease, GWA studies have not been successful in explaining the phenotypic heritability of age-associated chronic conditions[60] Alzheimer's, cancer and hypertension, indicating a polygenic role in disease association[61]. These genetic markers have been known to have small effect size and increasing sample sizes of study population may not always be feasible or result in reproducible set of variants[62]. Complex diseases are also affected by environmental factors, which may alter biological activity without completing inhibiting the functional consequence of the genes[63]. Therefore, an in-depth investigation of

genetic-variant influences on functional association may shed much-needed light on the age-related pathophysiology of chronic conditions and their co-prevalence to better therapeutic design and clinical trial cohort design[64].

GLOBAL HYPOTHESIS: Novel genetic risk factors underlie the observed patterns of co-occurrence in hypertension, cancer (and cancer therapy response), and Alzheimer's disease.

OVERALL APPROACH: As this study aims to investigate multiple disease associations our approach is to employ appropriate association analysis for unidirectional and divergent comorbidity patterns. Each association analysis is followed by integration of RNA-Sequence data, protein functional data and network analysis to identify multi-level biological changes influenced by individual's genetic profile available from public repositories (dbGaP, ADNI etc).

1.5 RESEARCH STRATEGY AND METHODOLOGY

1.5.1 Datasets

The population data for studying co-presence of complex disorders have been obtained via authorized access application for Alzheimer's Disease Neuroimaging Initiative (ADNI), and NCBI's Database of Genotype and Phenotype (dbGaP) for two datasets: 1) Breast and Prostate Cancer Cohort Consortium (BPC3) and 2) Genetic Predictors of Adverse Radiotherapy (Gene-PARE).

The **Gene-PARE study sample** was split into a discovery set (N=367) and a replication set (N=417). The phenotypes available with Gene-PARE are Subject ID, ethnicity, erectile dysfunction, prostatic rectal bleeding, AUASS (American Urological Association Symptom Score) score before and after radiotherapy, prostate resection, age at time of radiotherapy, radiotherapy type, anti-androgen therapy, prostate tumor stage, Gleason score, PSA, CT scan,

radiation dose, smoking, diabetes, and hypertension status, and use of alpha-blocker drugs among participants with or without prostate cancer. The Gene-PARE discovery dataset was provided in Affymetrix's SNP 6 array files (*.CEL files), therefore genotypes were called using bird-seed v2 algorithm, with thresholds of 97% call rate and HWE. Due to the absence of the reference files, R scripts were written and optimized to transform the individual genotype files for its use in Plink as .tped and .tfam files.

The **Breast and Prostate Cancer Cohort Consortium (BPC3)** was established in 2003 to conduct research on gene-environment interactions in cancer etiology. The sample size for prostate cancer individuals is 4069 and breast cancer is 595. The phenotypes available for BPC3 are Subject ID, affection status - breast or prostate cancer, age, family history of cancer and ancestry. The dataset is genotyped on Human660W-Quad using hg18 build.

The **ADNI data** includes 758 individuals with genotype data, and phenotypes of age, sex, ancestry, APOE status, affection status- mild cognitive impairment (MCI) and/Alzheimer's Disease (AD) and years of education. The dataset is genotyped on Illumina Human610-Quad BeadChip using hg18 build.

After two-level QC filtering (explained later), the datasets – Gene-PARE, ADNI, Breast and Prostate cancer– had 305, 677, 578 and 3857 individuals, respectively. The genome build for Gene-PARE is hg19, and hg18 for ADNI and BPC3, SNPs were mapped within 10kb to ~15,400 to 18000 genes using appropriate build reference for annotation.

1.5.2 Quality Control Protocol

Due to the large number of genetic loci being tested in genome wide studies, it is recommended to maximize the remaining number of markers in the study. Therefore, removing individuals with missing informations prior to removing to markers is a recommended quality-control approach. The quality control protocol was adapted from Nature Protocol by Anderson et al[65] and Clark et al[66] and the steps are outlined in Figure 9.

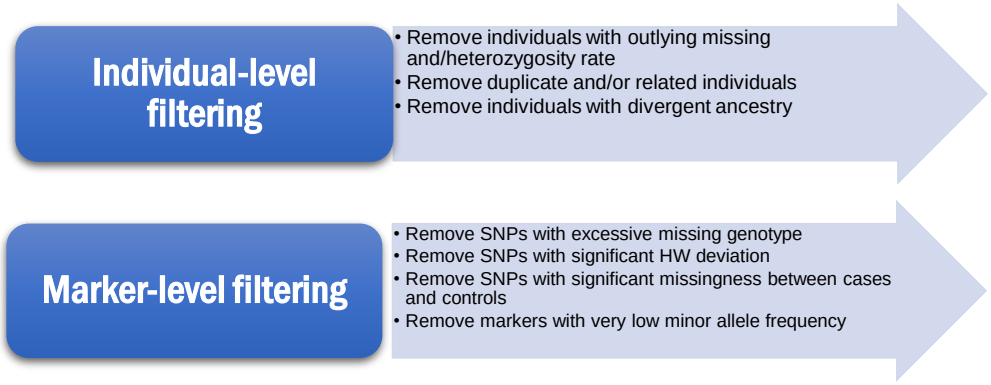


Figure 9: Steps for two-level quality control approach of genotype data. (Summarized from Anderson et al.)

1.5.3 Research Approach

1.5.3.1 Rationale

The genome wide association studies have been successful in discovering genetic variants for mendelian disorders and few replicable genetic loci for complex disease[55] genetic variants. However, due to multifactorial etiology of complex diseases, SNP-based GWA studies, despite increase in sample sizes, explain only a portion of phenotypic heritability. This approach has guided our attention towards polygenic influence in complex diseases, therefore due to small effect sizes of individual genetic loci, SNP-based GWAS are under powered to detect disease associations when genetic markers are correlated and high in number. Gene-based or region-based association testing have comparatively more power to detect gene-disease relationship as it aggregates SNPs into genes to test for joint association of markers with phenotype. Several gene-

based association studies have found phenotypic-associations which were missed in SNP-based GWAS, and later confirmed by meta-analysis of different cohorts, indicating the importance of gene-based analysis over single-marker-based association testing[67]. Complex disease genetics requires understanding the effect of SNPs on gene expression and functional products, and their association with disease phenotype. Therefore, pathway enrichment analysis combines genetic markers based on their biological properties, supplementing genotype-association testing in small sample case-control studies[68]. Elucidating the functional interaction affected from genetic variants requires integration of protein network data and tissue-specific expression activity [69]. Genetic variants have been proven to affect gene expression in multiple tissues, therefore integrating transcriptomic data with individual's genetic profile can assist in constructing a comprehensive understanding of multilevel biological changes caused due to complex diseases[70].

1.5.3.2 Overall strategy for genetic data integration

Since, this research study is investigating co-presence of complex diseases, the integrative pipeline to investigate genetic and trans-criptional perturbations of disease genes and its products in multiple tissues to understand in-teraction network of

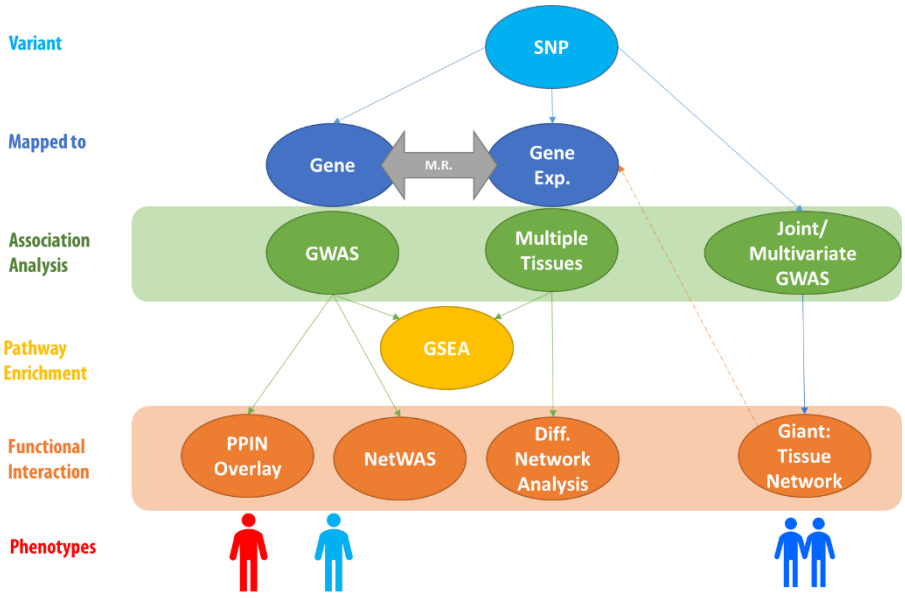


Figure 10: Proposed pipeline for integrative genomics data analysis strategy. (M.R. – Mendelian Randomization; GSEA- Gene Set Enrichment Analysis; PPIN- Protein-Protein Interaction Network; NetWAS – Network guided GWAS analysis; Giant – Genome-scale Integrated Analysis of Gene Network in Tissues. The icons in the phenotypes category represent individual analyses of the phenotypes, and the two figures together represent pooled phenotypes for multivariate analyses.)

disease genes is outlined in Figure 10. The genotype (SNP) data, is mapped to genes within 10kb window, to conduct a gene-based genome wide association study to identify contributing disease genes in individual phenotypes, the results are observed in a functional context using Network-assisted GWAS[71] (NetWAS) and protein-protein interaction overlay (PPIN). The gene expression of cis-SNPs is imputed from the genotype data to perform a gene-expression association study in approximately 30-40 tissue types using Genotype-Tissue Expression dataset (GTEx)[72], followed by conducting a differential gene network analysis. Mendelian Randomization (M.R.)[73] will be performed to investigate the direction of causality between genes, gene expression and phenotype. Additionally, the multivariate-GWAS will be conducted to test the effect of each SNP variant, all phenotypes under investigation are pooled together against common controls to identify the direction of odds ratio using multinomial regression, and marginal likelihood ratio from Bayesian model selection to identify cross-phenotypic effects of the diseases, the results of which will be translated with tissue-specific networks of gene interaction (GIANT)[71], and revalidated with gene expression association study from other data sources. The Joint-GWAS[74] will help in identification of diseases with inverse relationship, in a case-case GWA test by reducing background noise of high number of genetic markers[75].

1.5.3.3 Aim-specific association analysis

For **SPECIFIC AIM 1**, the side-effects of radiation therapy Erectile Dysfunction, Proctitis and Urinary Morbidity with IPSS/AUASS score monitered for 4 years after radiation therapy - will be analysed using direct and indirect method of multivariate GWAS. These traits have previously been investigated using single-phenotype-gene-variant association study for each phenotype. This study aims to increase the power of a smaller sample by analysing the adverse-effect phenotypes in a joint model using **1.a)** multinomial regression (direct method) and **1.b)** canonical corelation

analysis (indirect method) by reducing the number of tests and multiple testing burden. The multivariate approach will help in identification of genetic variants associated with multiple set of traits, to test for presence of pleiotropy.

For **SPECIFIC AIM 2**, the impact of hypertension will be tested using gene-based association analysis in each cohort to identify parallel genetic signals, if present. Additionally, comorbidity analysis is employed to test for association between hypertension-only diagnosis with individuals of Alzheimer's diagnosis against the history of hypertension (Alz+ vs Hyp-/ Alz+ vs Hyp+).

For **SPECIFIC AIM 3**, the inverse relationship is characterized using **3.a)** separate gene-based GWAS for breast and prostate cancer, and Alzheimer's disease to identify any changes in biological activity or off-target effects from tissue-specific network. For detecting **3.b)** directional-effect of SNPs, the controls are pooled and test against each phenotype (Alz+//B.CA+//P.CA+) using frequentist approach - multinomial regression analysis and Bayesian Model Selection, followed by to identify/negate influences of ascertainment or survival bias.

The association analysis is different for each aim, followed by imputation of gene expression profiles using reference transcriptomic data in multiple-tissue types, and common application of pathway enrichment using Gene-Set Enrichment Analysis (GSEA). The functional interaction will also be investigated in all association analysis using protein-protein interaction data integrated from STRINGdb[76] and NetWAS will be applied to identify tissue-based network priortized associations. The imputed gene expression data will be investigated using differential network analysis to identify network changes between cases and controls[77].The genetic variants found to be significant in Joint/multivariate GWAS will be priortized using GIANT-- multigene based query facilitated by tissue-based networks.

1.5.4 Analytical Tools

(Disclaimer: Description provided for the tools/software are included from excerpts from primary literature source)

Plink[78] is a C/C++ command based tool, and is used for conducting quality control procedure and generating binary bim/bed/fam files for association testing.

EIGENSTRAT[79] is used for principal component analysis for controlling population stratification.

MAGMA[80] is a software implementing multiple linear principal components regression to account for gene size and LD, and uses permutation up to 1,000,000 times to correct for multiple-testing. Their model projects the SNP matrix for a gene onto its principal components (PC), trimming away PCs with very small eigenvalues, and then uses those PCs as predictors for the phenotype in the linear regression model. This improves power by removing redundant parameters, and guarantees that the model is identifiable in the presence of highly collinear SNPs. It also allows covariate adjustment for association testing.

PrediXcan[70] predicts gene expression as a function of genetic variants using a reference dataset such as GTEx[72], where individuals have been genotyped for variants and expression profiles using elastic net model.

Trinculo[81] is a open source C based command line program using plink binary files as input and allows covariate adjustment for conducting frequentist – multinomial regression, and bayesian approach – Bayesian Model Selection to identify multi-trait sub-phenotypes and cross-disorder association studies.

MV-Plink[82] stands for multivariate-Plink, and uses Canonical Correlation Analysis (CCA) to identify a linear combination of multiple traits producing a F-test and p-value for each variant considered.

**B.CA – Breast Cancer; P.CA – Prostate Cancer; Hyp- Hypertension; + → presence; - → absence*

functional charcateristics using hypergeometric test against gene sets obtained from different functional data repositories- MSigDB, Gene Ontology, Reactome and WikiPathways.

STRINGdb[76] is Search Tool for the Retrieval of Interacting Genes/Proteins, a web resource for known and predicted protein-protein, protein-DNA interaction network.

NetWAS[71] combines genes with nominally significant genome-wide association study (GWAS) P values and tissue-specific networks to identify disease-gene associations.

CytoDDN[77] is a cytoscape-plugin for analysis of differential dependency network (DDN) for detection and visualization of statistically significant topological changes in transcriptional networks representing two biological conditions.

GIANT[71] provides an interface to human tissue networks through multi-gene queries, network visualization enableing systematic exploration of the landscape of interacting genes that shape specialized cellular functions across more than a hundred human tissues and cell types.

RStudio[84] is a free and open-source integrated development environment (IDE) for R -- a programming language for statistical computing and graphics.

1.5.5 Alternative Strategies

To minimize low power of small sample sets, the approach of the study is to combine multiple data to prioritize the gene variant association, aggregating them based on region, functional activity and enrichment of set of markers with known pathways. However, it is possible that effect of

individual rare variants may not be detected, and therefore, the alternative is to utilize summary statistics from other GWA studies with bigger sample sizes to validate the associations with higher significance value ($p<0.01$). Additionally, associations with higher p-values can also be validated in datasets with genetic profile of a large sample size, however access to such large sample sizes is either not available or suffers from poor phenotype cataloguing. Mendelian randomization determines if there is a causal effect between a suspected risk factor and its associated outcome. Mendelian randomization rests on 3 assumptions: (1) the genetic variant is associated with the risk factor; (2) the genetic variant is not associated with confounders; and (3) the genetic variant influences the outcome only through the risk factor. The second and third assumptions are independent from biological pleiotropy, and since we testing under the hypothesis of multiple pleiotropic models, and if the hypothesis is failed to be rejected, then Mendelian Randomization at this juncture may prove to be biased[73]. Therefore, mendelian randomization approach to test for intermediate pleiotropy will be applied if the study fails to detect biological pleiotropic associations.

For imputation of gene expression profile using transcriptomic data, the alternative approach is to apply Slinger[85] which relies on unrestricted set of SNPs. At present, Slinger has been trained on genotype data from microarray technology, resulting in a large but still incomplete set of SNPs and lacks multiple tissue type prediction, but upcoming updates to this statistical tool would make this software a supplementary analytical approach to PrediXcan.

1.6 EXPECTED OUTCOMES

This research study takes advantage of the publicly available genotype data to investigate two major comorbidity patterns by integrating transcriptomic level information, and protein-protein

interaction network. Since comorbidity patterns are prevalent in the aging population, this is the first study to contribute towards information on shared pathologies and possible off-target effects of hypertension, cancer, and Alzheimer's disease. Additionally, the study will provide insights into multiple pleiotropic models that may be found to be associated with chronic conditions and can be used to suggest clinical decisions and management, screening measures, and polypharmacy concerns in the elderly population[12].

1.7 REFERENCES

1. Mark Mather, L.A.J., Kelvin M. Pollard. Aging in the United States. in *Population Bulletin* Vol. 70 (Population Reference Bureau, 2015).
2. Johnson, S.C., Dong, X., Vijg, J. & Suh, Y. Genetic evidence for common pathways in human age-related diseases. *Aging Cell* **14**, 809-17 (2015).
3. Alzheimer's, A. 2011 Alzheimer's disease facts and figures. *Alzheimers Dement* **7**, 208-44 (2011).
4. Noone AM, H.N., Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA SEER Cancer Statistics Review. in *Cancer Statistics Review* (National Cancer Institute, Bethesda, MD, 1975-2015).
5. Miller, K.D. *et al.* Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* **66**, 271-89 (2016).
6. Fryar, C.D., Ostchega, Y., Hales, C.M., Zhang, G. & Kruszon-Moran, D. Hypertension Prevalence and Control Among Adults: United States, 2015-2016. *NCHS Data Brief*, 1-8 (2017).
7. Salive, M.E. Multimorbidity in older adults. *Epidemiol Rev* **35**, 75-83 (2013).
8. Williams, G.R. *et al.* Comorbidity in older adults with cancer. *J Geriatr Oncol* **7**, 249-57 (2016).
9. Musicco, M. *et al.* Inverse occurrence of cancer and Alzheimer disease: a population-based incidence study. *Neurology* **81**(2013).
10. Capobianco, E. & Lio, P. Comorbidity: a multidimensional approach. *Trends Mol Med* **19**, 515-21 (2013).
11. Sharma, A. *et al.* Network-based analysis of genome wide association data provides novel candidate genes for lipid and lipoprotein traits. *Mol Cell Proteomics* **12**, 3398-408 (2013).
12. Nobili, A., Garattini, S. & Mannucci, P.M. Multiple diseases and polypharmacy in the elderly: challenges for the internist of the third millennium. *Journal of Comorbidity* **1**, 28-44 (2011).
13. Cheng, X. & Jin, V.X. An Introduction to Integrative Genomics and Systems Medicine in Cancer. *Genes* **9**, 37 (2018).
14. Lopez-Otin, C., Blasco, M.A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194-217 (2013).
15. Aunan, J.R., Watson, M.M., Hagland, H.R. & Soreide, K. Molecular and biological hallmarks of aging. *Br J Surg* **103**, e29-46 (2016).
16. Maman, S. & Witz, I.P. A history of exploring cancer in context. *Nature Reviews Cancer* (2018).
17. Gupta, G.P. & Massagué, J. Cancer Metastasis: Building a Framework. *Cell* **127**, 679-695 (2006).
18. Torre, L.A. *et al.* Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87-108 (2015).
19. Fargo, K. & Bleiler, L. Alzheimer's association report. *Alzheimers Dement* **10**(2014).
20. Jorgensen, T.L., Hallas, J., Friis, S. & Herrstedt, J. Comorbidity in elderly cancer patients in relation to overall and cancer-specific mortality. *Br J Cancer* **106**, 1353-60 (2012).
21. Hu, J.X., Thomas, C.E. & Brunak, S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* **17**, 615-29 (2016).

22. Services, C.f.M.a.M. Chronic Conditions Among Medicare Beneficiaries, Chartbook, 2012 Edition. in *Centers for Medicare and Medicaid Services* (Baltimore, MD, 2012).
23. Cheng, P. & Caram, M.V. Comorbid conditions in patients with castration-resistant prostate cancer (CRPC) receiving abiraterone. *Journal of Clinical Oncology* **34**, e16578-e16578 (2016).
24. Owen, J.B. *et al.* Can Patient Comorbidities Be Included in Clinical Performance Measures for Radiation Oncology? *Journal of Oncology Practice* **10**, e175-e181 (2014).
25. West, C.M. & Barnett, G.C. Genetics and genomics of radiotherapy toxicity: towards prediction. *Genome Med* **3**, 52 (2011).
26. Gomez-Millan, J. Radiation therapy in the elderly: more side effects and complications? *Crit Rev Oncol Hematol* **71**, 70-8 (2009).
27. Hjalm-Eriksson, M. *et al.* Comorbidity as a predictor of overall survival in prostate cancer patients treated with external beam radiotherapy combined with HDR brachytherapy boosts. *Acta Oncol* **56**, 21-26 (2017).
28. Sogaard, M., Thomsen, R.W., Bossen, K.S., Sorensen, H.T. & Norgaard, M. The impact of comorbidity on cancer survival: a review. *Clin Epidemiol* **5**, 3-29 (2013).
29. Duthie, A., Chew, D. & Soiza, R.L. Non-psychiatric comorbidity associated with Alzheimer's disease. *QJM* **104**, 913-20 (2011).
30. Haaksma, M.L. *et al.* Comorbidity and progression of late onset Alzheimer's disease: A systematic review. *PLOS ONE* **12**, e0177044 (2017).
31. Judd, E. & Calhoun, D.A. Apparent and true resistant hypertension: definition, prevalence and outcomes. *Journal of human hypertension* **28**, 463-468 (2014).
32. Wang, J.H., Wu, Y.J., Tee, B.L. & Lo, R.Y. Medical Comorbidity in Alzheimer's Disease: A Nested Case-Control Study. *J Alzheimers Dis* **63**, 773-781 (2018).
33. Lyall, D.M. *et al.* Associations between single and multiple cardiometabolic diseases and cognitive abilities in 474 129 UK Biobank participants. *Eur Heart J* **38**, 577-583 (2017).
34. Sery, O., Hlinecka, L., Balcar, V.J., Janout, V. & Povova, J. Diabetes, hypertension and stroke - does Alzheimer protect you? *Neuro Endocrinol Lett* **35**, 691-6 (2014).
35. Edwards, J.D. *et al.* Antihypertensive Treatment is associated with MRI-Derived Markers of Neurodegeneration and Impaired Cognition: A Propensity-Weighted Cohort Study. *J Alzheimers Dis* **59**, 1113-1122 (2017).
36. Feng, Y.A. *et al.* Investigating the genetic relationship between Alzheimer's disease and cancer using GWAS summary statistics. *Hum Genet* **136**, 1341-1351 (2017).
37. Roe, C.M., Behrens, M.I., Xiong, C., Miller, J.P. & Morris, J.C. Alzheimer disease and cancer. *Neurology* **64**(2005).
38. Roe, C.M. *et al.* Cancer linked to Alzheimer disease but not vascular dementia. *Neurology* **74**(2010).
39. White, R.S., Lipton, R.B., Hall, C.B. & Steinerman, J.R. Nonmelanoma skin cancer is associated with reduced Alzheimer disease risk. *Neurology* **80**(2013).
40. Freedman, D.M. *et al.* Associations between cancer and Alzheimer's disease in a U.S. Medicare population. *Cancer Medicine* **5**, 2965-2976 (2016).
41. Bowles, E.J.A. *et al.* Risk of Alzheimer's disease or dementia following a cancer diagnosis. *PLOS ONE* **12**, e0179857 (2017).
42. Akushevich, I. *et al.* Morbidity risks among older adults with pre-existing age-related diseases. *Experimental Gerontology* **48**, 1395-1401 (2013).

43. Frain, L. *et al.* Association of cancer and Alzheimer's disease risk in a national cohort of veterans. *Alzheimers Dement* **13**, 1364-1370 (2017).
44. Ma, L.L. *et al.* Association between cancer and Alzheimer's disease: systematic review and meta-analysis. *J Alzheimers Dis* **42**, 565-73 (2014).
45. Becker, K. *et al.* Meta-Analysis of Genome-Wide Association Studies and Network Analysis-Based Integration with Gene Expression Data Identify New Suggestive Loci and Unravel a Wnt-Centric Network Associated with Dupuytren's Disease. *PLoS One* **11**, e0158101 (2016).
46. Williams, C. *et al.* Transcriptome analysis of synaptoneurosomes identifies neuroplasticity genes overexpressed in incipient Alzheimer's disease. *PLoS One* **4**(2009).
47. Behrens, M.I., Lendon, C. & Roe, C.M. A common biological mechanism in cancer and Alzheimer's disease? *Curr Alzheimer Res* **6**(2009).
48. Ibáñez, K., Boullosa, C., Tabarés-Seisdedos, R., Baudot, A. & Valencia, A. Molecular Evidence for the Inverse Comorbidity between Central Nervous System Disorders and Cancers Detected by Transcriptomic Meta-analyses. *PLOS Genetics* **10**, e1004173 (2014).
49. Sanchez-Valle, J. *et al.* A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's Disease, Glioblastoma and Lung cancer. *Sci Rep* **7**, 4474 (2017).
50. Driver, J.A. Inverse association between cancer and neurodegenerative disease: review of the epidemiologic and biological evidence. *Biogerontology* **15**, 547-57 (2014).
51. Rubio-Perez, C. *et al.* Genetic and functional characterization of disease associations explains comorbidity. *Sci Rep* **7**, 6207 (2017).
52. Chen, Y. & Xu, R. Network Analysis of Human Disease Comorbidity Patterns Based on Large-Scale Data Mining. in *Bioinformatics Research and Applications* 243-254 (2014).
53. Klimek, P., Aichberger, S. & Thurner, S. Disentangling genetic and environmental risk factors for individual diseases from multiplex comorbidity networks. *Scientific Reports* **6**, 39658 (2016).
54. American Geriatrics Society Expert Panel on the Care of Older Adults with, M. Guiding Principles for the Care of Older Adults with Multimorbidity: An Approach for Clinicians. *Journal of the American Geriatrics Society* **60**, E1-E25 (2012).
55. Fridley, B.L. & Biernacka, J.M. Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur J Hum Genet* **19**, 837-43 (2011).
56. Nudelman, K.N. *et al.* Association of cancer history with Alzheimer's disease onset and structural brain changes. *Front Physiol* **5**, 423 (2014).
57. Koestler, D.C., Jones, M. & Kobor, M. The era of integrative genomics: more data or better methods? *Epigenomics* **6**, 463-467 (2014).
58. Wu, Y. *et al.* Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun* **9**, 918 (2018).
59. Lin, J.R. *et al.* Integrated Post-GWAS Analysis Sheds New Light on the Disease Mechanisms of Schizophrenia. *Genetics* **204**, 1587-1600 (2016).
60. Rappaport, S.M. Genetic Factors Are Not the Major Causes of Chronic Diseases. *PLoS ONE* **11**, e0154387 (2016).
61. Abraham, G. & Inouye, M. Genomic risk prediction of complex human disease and its clinical application. *Current Opinion in Genetics & Development* **33**, 10-16 (2015).
62. Visscher, P.M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* **101**, 5-22 (2017).

63. Pranavchand, R. & Reddy, B.M. Genomics era and complex disorders: Implications of GWAS with special reference to coronary artery disease, type 2 diabetes mellitus, and cancers. *Journal of Postgraduate Medicine* **62**, 188-198 (2016).
64. Bebek, G., Koyutürk, M., Price, N.D. & Chance, M.R. Network biology methods integrating biological data for translational science. *Briefings in Bioinformatics* **13**, 446-459 (2012).
65. Anderson, C.A. *et al.* Data quality control in genetic case-control association studies. *Nature Protocols* **5**, 1564 (2010).
66. Clarke, G.M. *et al.* Basic statistical analysis in genetic case-control studies. *Nature Protocols* **6**, 121 (2011).
67. Peng, G. *et al.* Gene and pathway-based second-wave analysis of genome-wide association studies. *European Journal Of Human Genetics* **18**, 111 (2009).
68. Jin, L. *et al.* Pathway-based analysis tools for complex diseases: a review. *Genomics Proteomics Bioinformatics* **12**, 210-20 (2014).
69. Cline, M.S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2**, 2366-82 (2007).
70. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091-1098 (2015).
71. Greene, C.S. *et al.* Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet* **47**, 569-76 (2015).
72. Consortium, G.T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
73. Emdin, C.A., Khera, A.V. & Kathiresan, S. Mendelian randomization. *JAMA* **318**, 1925-1926 (2017).
74. McGeachie, M.J. *et al.* Joint GWAS Analysis: Comparing similar GWAS at different genomic resolutions identifies novel pathway associations with six complex diseases. *Genom Data* **2**, 202-211 (2014).
75. Curtis, D. *et al.* Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes. *Psychiatr Genet* **21**, 1-4 (2011).
76. Jensen, L.J. *et al.* STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412-6 (2009).
77. Zhang, B. *et al.* DDN: a caBIG(R) analytical tool for differential network analysis. *Bioinformatics* **27**, 1036-8 (2011).
78. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
79. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904 (2006).
80. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology* **11**, e1004219 (2015).
81. Jostins, L. & McVean, G. Trinculo: Bayesian and frequentist multinomial logistic regression for genome-wide association studies of multi-category phenotypes. *Bioinformatics* **32**, 1898-900 (2016).
82. Ferreira, M.A. & Purcell, S.M. A multivariate test of association. *Bioinformatics* **25**, 132-3 (2009).

83. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
84. RStudioTeam. RStudio: Integrated Development for R. RStudio, Inc., . (Boston, MA, 2015).
85. Vervier, K. & Michaelson, J.J. SLINGER: large-scale learning for predicting gene expression. *Sci Rep* **6**, 39360 (2016).
86. Teschendorff, A.E. & Relton, C.L. Statistical and integrative system-level analysis of DNA methylation data. *Nat Rev Genet* **19**, 129-147 (2018).
87. Galanter, J.M. *et al.* Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *eLife* **6**, e20532 (2017).
88. Abi Aad, S. *et al.* Hypertension induced by chemotherapeutic and immunosuppresive agents: A new challenge. *Critical Reviews in Oncology / Hematology* **93**, 28-35 (2015).

Chapter 2

SNP-DERIVED TRANSCRIPTOMICS & COPY NUMBER VARIATION IN PROCTITIS— A RADIOTHERAPY SIDE EFFECT—POINTS TO ALTERED MITOCHONDRIAL AND DNA REPAIR MECHANISMS

Gita A Pathak, Nicole R Phillips

Submitted to Scientific Reports (June 2019)

ABSTRACT. Proctitis is an inflammation of the rectum and may be induced by radiation treatment for cancer. We investigated proctitis as a radiotoxic endpoint in prostate cancer patients who received radiotherapy (n=222). We analyzed the copy number variation and SNP-derived transcriptomic profiles of whole-blood and prostate tissue associated with proctitis. The SNP and copy number data were genotyped on Affymetrix® Genome-wide Human SNP Array 6.0. Following QC measures, the genotypes were used to obtain gene expression by leveraging GTEx, a reference dataset for gene expression association based on genotype and RNA-seq information for prostate (n= 132) and whole-blood tissue (n=369). In prostate tissue, 62 genes were significantly associated with proctitis, and 98 genes in whole-blood tissue. Six genes - *CABLES2*, *ATP6AP1L*, *IFIT5*, *ATRIP*, *TELO2*, and *PARD6G* were common to both tissues. The copy number analysis identified seven regions

associated with proctitis, one of which (*ALG1L2)* was also associated with proctitis based on transcriptomic profiles in the whole-blood tissue. The genes identified via transcriptomics and copy number variation association were further investigated for enriched pathways and gene ontology. Some of the enriched processes were DNA repair, mitochondrial apoptosis regulation, cell-to-cell signaling interaction processes for renal and urological system, and organismal injury.

Short Title: Genetic characterization of radiotherapy side effects

2.1 INTRODUCTION

Prostate cancer is one of the most prevalent diseases in older men, with 66 years being the average age at the time of cancer diagnosis[1]. According to the cancer statistics of 2019, approximately 3 million men have been previously diagnosed with prostate cancer and are still alive today. This feat can be credited to the advancement in cancer treatment which has contributed to the 5-year relative survival rate of 90% in prostate cancer survivors[2]. Radiation therapy is one of the primary forms of treatment for prostate cancer, delivered as external beam radiotherapy (EBRT) or brachytherapy. While the dose and precision of radiation delivery to the tumor tissue has improved over the years, surrounding normal tissue get irradiated leading to clinical side effects [3]. Proctitis is the inflammation of the rectum, which can result from receiving radiation therapy around the pelvic region such as in prostate cancer treatment[4]. The inflammation of the rectum can either be acute or chronic. Acute proctitis appears within 3 months of receiving radiation therapy, and progression of rectal inflammation after 3 months of completing radiation therapy is identified as chronic proctitis [5]. The development of radiation-induced chronic proctitis affects 5-20% of cancer survivors and is relatively more common[6] than acute proctitis which affects approximately 13% of the cancer population[5]. As of 2016, the population of cancer survivors in the US was estimated to be approximately 15 million, and by the year 2026 is expected to reach 20 million individuals [7]. Given the prevalence of chronic proctitis affecting cancer individuals (5-20%), we can deduce that approximately 1-4 million cancer survivors experience proctitis from receiving radiotherapy. The goal of radiation therapy in treating cancer is to damage the DNA of cancer tissue by creating double-strand breaks (DSBs). While the cellular system is capable of repairing breaks in the DNA, strands with DSBs are difficult to restore leading to activation of apoptotic signals and ultimately killing cancer cells. Unfortunately, the normal tissue around the targeted region is also affected by

DNA damage from radiation, and have to rely on DNA repair mechanisms for rehabilitation of cellular functions [8]. Recent twin-study has reported that certain SNPs and their transcriptomic influence is associated with individual radiation sensitivity and a heritability estimate of 66% [9]. Therefore, it is vital to understand genetics underlying molecular mechanisms involved in adverse effects of radiotherapy and individual genetic variations that may induce radiation sensitivity. Genome-wide studies have been conducted to identify gene variants that may contribute towards developing radiotoxic side-effects. These studies have identified genetic variations involved in DNA repair pathways to be associated with overall radiotoxicity[3,10]. However, the role of altered gene expression[9] from aggregated single nucleotide polymorphisms (SNPs) remains elusive in radiotoxicity phenotypes (e.g. proctitis). Regulatory variants are SNPs within coding regions which contribute towards tissue – specific gene expression alterations leading to wide variations in the phenotypic spectrum. Estimating the contribution of SNP aggregates to gene expression can be carried out using correlation weights derived from reference datasets which contain both SNP and RNA-seq information as modelled in PrediXcan[11]. One such dataset is the GTEx project, an NIH funded initiative that stores genotype and RNA-seq data of 53 tissues from 620 donors (v7). The majority of the donors in the GTEx dataset are Caucasian, and more than 50% of the donors are over the age of 50 years[12]. These characteristics make GTEx an excellent reference dataset to derive gene expression values from individual level SNP profiles of prostate cancer patients who have received radiation treatment.

Beyond SNPs, genetic discordance from gene dosage and structural effects can be attributed to copy number variation (CNV). CNVs are segments of DNA that are greater than 1kb with differences in size between the two copies[13]. In a clinical setting, testing of CNVs is relatively more common than other genetic tests[14] due to the majority of phenotypic changes associated with

these variations in segment size. CNVs associated with radiotoxicity phenotypes (e.g. proctitis) have not been investigated extensively and may prove to be significant contributors to phenotype risk.

We hypothesize that genotype-derived gene expression profiles and variations in copy number will identify genetic alterations associated with a spectrum of DNA repair functions. Here, we investigate both CNV and tissue specific (prostate and whole-blood) transcriptomic profiles derived from individual-level SNPs that are associated with radiation induced proctitis in prostate cancer patients.

2.2 MATERIAL & METHODS

The overall methodology is visually summarized in Fig S1.

Data access to study subjects. The Gene-PARE was approved by the Institutional Review Board of the Icahn School of Medicine at Mount Sinai and Florida Radiation Oncology Group (Kerns et al.[3,15]). All patients provided informed consent under the parent study – Gene-PARE, at Icahn School of Medicine, Mount Sinai and Florida Radiation Oncology Group (Kerns et al.[3,15]). We obtained access to anonymized individual level genotype data from Genetic Predictors of Adverse Radiotherapy Effects (Gene-PARE) (phs000772.v1.p1) via dbGaP's authorized application (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000772.v1.p1&phv=202663&phd=&pha=&pht=3996&phvf=&phdf=&phaf=&phtf=&dssp=1&consent=&temp=1) under the approval of North Texas Regional IRB protocol 2016-090. The study described here was performed under the North Texas Regional IRB (formerly the University of North Texas Health Science Center IRB), and was given "EXEMPT" status based on the criteria that our study involved data available from public repository, i.e. dbGaP

and does not require approval of receiving informed consent. We analyzed prostate cancer individuals from the discovery set (N=367), which were genotyped for 934,940 SNPs on Affymetrix® Genome-wide Human SNP Array 6.0. The dataset contains phenotypic information on prostate cancer patients who have received radiation treatment either via EBRT or brachytherapy. Out of three radiotoxicity phenotypes – erectile dysfunction, proctitis and urinary morbidity (IPSS/AUASS) – we focused our investigation on proctitis because (1) it is also prevalent in other pelvic region cancers[16] and (2) the dataset for proctitis was complete for all individuals.

SNP-QC. We extracted SNP data from the *.CEL files using Affymetrix® Genotyping Console using the BIRDSEED v2 algorithm and genotype call rate of 95% and default settings from the array, leaving 905,280 markers and total of 355 individuals. The files were then exported to plink[17] format to perform QC measures as suggested by Anderson et. al[18]. At the individual-level filtering, we removed 5 individuals for either failing heterozygosity or having greater than 9% missing genotypes. At the IBD filter of 0.1875, we removed 29 individuals; further, 120 individuals were removed who were not Caucasian or failed to cluster with the main patient population based on principal component (PC) analysis of PC1 and PC2. After SNP-level filtering on SNP missingness, minimum allele frequency and Hardy-Weinberg equilibrium, we were left with 746,684 SNPs and 222 individuals. The final cohort characteristics after QC were analyzed using Fischer's exact test for categorical variables and student t-test for continuous variables (Table 1).

Table1: *Characteristics of individuals with prostate cancer who received radiotherapy.*

	Prostate cancer individuals without proctitis; Controls *(N= 177)*	**Prostate cancer individuals with proctitis; Cases *(N=45)***	***P-value***
Mean ± SD			
Age	63.7 ± 7.35	65. 28 ± 7.69	0.25
Gleason score	6.45 ± 0.78	6.31 ± 0.72	0.26
N (%)			
Smoking			
Yes	67 (38%)	14 (31%)	0.48
No	110 (62%)	31 (69%)	
Diabetes			
Yes	5 (3%)	3 (7%)	0.21
No	172 (97%)	42 (93%)	
Hypertension			
Yes	55 (31%)	10 (22%)	0.28
No	122 (69%)	35 (78%)	
Treatment			
EBRT & Brachytherapy	73 (41%)	21 (47%)	0.15
EBRT	0 (0%)	1 (2%)	
Brachytherapy	104 (59%)	23 (51%)	

Gene Expression imputation and GSEA. Tissue specific gene expression prediction using individual's genotype profile was performed using PrediXcan[11]. The weights of SNPs and tissue specific genes were trained using lasso regression and GTEx (v7)[12] reference datasets, accessible at http://predictdb.org/. We downloaded model files for prostate and whole-blood tissues. In the GTEx (v7), there are 132 prostate tissue donors and 369 donors for whole-blood tissue. PrediXcan implements gene expression value prediction, followed by gene-based association. The z-scores identify the direction of expression[19] for each genes and their corresponding p-values for association testing. For prostate tissue, 3113 genes were predicted and 5954 genes for whole-blood tissue. Following association tests, significant genes (identified as p-value <0.05) were investigated further by constructing tissue-specific protein-protein interaction (PPI) networks between query (significant genes) and interacting genes using DifferentialNet[20] database and NetworkAnalyst3.0 [21]. The network was filtered on betweenness centrality of 4.0 in order to reduce isolated neighboring nodes (each gene is a node). All the genes in the network were subsequently

analyzed for gene set enrichment using clusterProfiler[22] for gene ontology and visualized in GOPlot[23].

CNV association and GSEA. The *.CEL files of 222 individuals from the above QC protocol were extracted for copy number analysis using Affymetrix® Genotyping Console. Copy number segments were filtered to regions (minimum genomic size of 2kbps) with 10 marker per segment[24]. The copy number data was exported as tab-delimited file for copy number association in CNVRuler[25]. CNV regions were considered to be significantly associated at FDR p-value <0.05. The significant CNV regions were visualized using Phenogram[26] then mapped to genes using UCSC browser for GRCh37/hg19 assembly (https://genome.ucsc.edu/). The genes within CNV regions were analyzed for functional and diseases processes and visualized using Ingenuity Pathway Analysis® (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis).

2.3 RESULTS

Genes identified in prostate tissue. In the association analysis between prostate cancer individuals who developed proctitis (cases) and who didn't develop proctitis (controls), we found a total of 62 differentially expressed genes to be significantly associated. Based on z-score direction, 28 genes were downregulated, and 34 genes were upregulated in the prostate tissue (Table S1). We mapped the genes to tissue-specific protein-protein interaction (PPI) network to understand combined functional effects of differentially expressed genes followed by analyzing all the genes in the network for enriched gene ontology of biological processes, molecular functions and cellular components (Fig 1). Some of the key processes and their contributing genes identified, were protein deubiquitination (*ARRB2, TP53, SHMT2, BRCA1, ESR1, NEDD8, MYC*),

wnt signaling (*MOV10, ARRB2, LRRK2, TNIK, ESR1, APP, CUL3*) , regulation of apoptosis signaling (*ARRB2, TP53, LRRK2, BRCA1, YWHAZ, PTTG1IP*), response to radiation (*CIRBP, TP53, BRCA1, APP, MYC*) and mitochondrial organization & apoptotic mitochondrial changes (*ARRB2, TP53, LRRK2, YWHAZ*) (Table S2).

Genes identified in whole blood tissue. We found a total of 98 genes to be associated with proctitis in whole blood tissue. 49 genes were upregulated, and 49 genes were downregulated (Table S3). Integrating PPI network information with the significant genes, highlighted DNA repair processes (Fig 2) such as DNA replication (*TERF2, EGFR, CDC7, BRCA1, ATRIP, RBBP8, SLX4, ORC1, ORC6, RAD50, CDK2, MCM2, DTL, RPA1, RPA2, RPA3*), DNA integrity checkpoint (*FBXO6, BRCA1, CDC5L, ATRIP, MDM2, ORC1, FZR1, CDK2, TP53, DTL, RPA2*), nucleotide excision repair (*COPS6, RBBP8, DDB1, UBC, SLX4, TP53, RPA1, RPA2, RPA3*), recombinational repair (*CDC7, BRCA1, RBBP8, SLX4, RAD50, RPA1, RPA2, RPA3*), detection of DNA damage and response (*DDB1,UBC,DTL,RPA1,RPA2,RPA3*) and telomeric maintenance (*TERF2, CCT5, SLX4, RAD50, TELO2, RPA1, RPA2, RPA3*) (Table S4).

CNV association and GSEA of mapped genes. We found 7 CNV regions associated with proctitis on chromosomes 1, 3, 4, 11, 12 and 15 (Table S5). We identified genes within CNV regions using UCSC browser (hg19) (Table S6). Interestingly, out of the two regions on chromosome 11 that were significant, we observe a high number of TRIM family genes (chr11:89487937-89909274 bps). The mapped genes from copy number regions were investigated for gene interactions using biobase knowledge of Ingenuity Pathway Analysis®. The pathway with the highest number of query genes (Fig 4) was further analyzed for enriched disease and functional categories (Table S7). Cell-to-cell signaling interaction processes for renal and urological system, connective tissue development and function, and organismal injury were significantly associated

processes, and their functional categorization included synthesis, proliferation, apoptosis and transmembrane transport. It is interesting to note, that most of these processes were dominated by *TRH* and *TRIM*-family genes. Furthermore, we also observed that the *ALG1L2* gene, which was one of the significantly downregulated genes in whole-blood tissue, was also mapped to significant CNV region on chr3:129690192-129896364 bps which observes both gain and loss of copy, referred to as mixed regions.

2.4 DISCUSSION

Genetic susceptibility towards developing radiotoxic phenotypes is an upcoming research interest of significant clinical impact to improve the quality of life of cancer survivors[8]. Previously, GWAS studies have been conducted to identify genetic loci associated with overall toxicity, decreased urinary stream, and erectile dysfunction[10,27] in prostate cancer individuals who received radiotherapy[28]. While these findings have shed much-needed light on SNP loci associated with susceptibility towards radiotoxicity, cumulative effects of exonic SNPs on gene expression and other genetic alterations such as copy number differences have not been previously studied. Here, we integrated genotyping data to identify genetic risk associated with proctitis by (1) employing genetic variant–derived gene expression of both prostate and whole-blood tissue, and (2) identifying associated genomic CNVs. The transcriptomic analyses points to several novel genes that play role in DNA-repair processes. In addition, we identified variable copy number regions had multiple members of TRIM-family genes to be associated with proctitis. Along with novel genes identified through the analysis, the incorporation of PPI map reveals convergence of the implicated gene sets on known DNA-repair, mitochondrial, and telomeric regulation processes, highlighting their involvement with radiotoxic phenotypes (e.g. proctitis).

Six genes from both prostate and whole-blood tissue were associated with proctitis. *CABLES2*, *ATP6AP1L and IFIT5* were under expressed, and *ATRIP* and *TELO2* were upregulated in both tissues, however *PARD6G* was over expressed in prostate tissue and under expressed in whole blood tissue. *CABLES2* (Cdk5 And Abl Enzyme Substrate 2), which is involved in regulation of the cell cycle, was also reported to be under expressed in lymphocytes of occupational workers who were exposed to ionizing radiation[29]. *ATP6AP1L* (ATPase H+ Transporting Accessory Protein 1 Like) is critical for proton transportation and ATP synthase activity in the mitochondria; it has a paralog, ATP6AP1, which is involved in secretory granules and regulating neuroendocrine responses[30]. *IFIT5* (IFN-induced protein with tetratricopeptide repeats) has been reported to act as an enhancer in immune responses, with partial containment in mitochondria[31]. A recent study has reported that elevated *IFIT5* gene expression was correlated with interferon-γ levels in prostate cancer individuals after radiation, and demonstrated that IFN-γ stimulated epithelial-to-mesenchymal transition through the activation of JAK-STAT pathway[32]. *ATRIP* [*TREX1*] is an ATR interacting protein that is a DNA exonuclease[33] that is known to initiate DNA repair pathway[34]. *ATRIP* has also been reported to provide telomere protection by recruiting *ATM* – a key player in regulating cellular damage responses— to telomeric and DNA damage sites unaided by *ATR* kinase activity[35]. Additionally, *ATR* responds to UV damage via the downregulation of Pin1 demonstrating anti-apoptotic activity in mitochondria[36]. *ATRIP* [*TREX1*] has been shown to be upregulated in radiation-induced immunogenicity of tumor cells[37] by degrading cytosolic dsDNA and transferring cancer cells to dendritic cells under the stimulation of interferon-type1[38]. In addition to *ATRIP*, *TELO2* [*CLK2*] also interacts with ATM to stimulate cell cycle arrest in response to radiation induced double strand breaks[39] via AKT activation[40]. High *TELO2* expression activity has been identified to be correlated with cell protection when exposed to high

radiation doses[40]; conversely, *TELO2* overexpression can trigger inflammation by influencing PIKKs (via *mTORC1* binding[41]) while responding to DNA damage[42]. *PARD6G* was found to have opposite direction of expression in prostate and whole-blood tissue, which could be attributed to tissue specific differences. Hypermethylation and downregulation of PARD6G was concluded to be involved in DNA repair mechanisms[43] in bisphenol A (BPA, a xenoestrogen) exposed human-derived breast cancer epithelial cells.

Assessment of copy number variation associated with radiation toxicity phenotypes can help identify genetic alterations that may lead to functional changes in gene expression, and thus, phenotypic variation. While CNVs have recently gathered interest in cancer diagnosis and treatment, their importance in studying radiation toxicity phenotypes remains understudied. Here, analysis of copy number data identified 7 regions to be associated with proctitis. The *ALG1L2* mapped to the significant CNV region on chr3:129690192-129896364 (along with *TRH* and *FAM86HP)* was also found to be associated with proctitis in the transcriptomic analysis of whole-blood tissue, while the other two genes (*TRH* & *FAM86HP*) were not predicted in either of the two tissues. Based on gene ontology annotations, *ALG1L2* functions as a mannosyl transferase in protein glycosylation[44]. Another gene within chromosome 3 CNV region is *TRH* (Thyrotropin releasing hormone); its function includes carbohydrate and amino acid metabolism, and it has been involved in endocrine system disorders, metabolic disease, and organismal injury (Table S7). *TRH* has been shown to mobilize calcium from endoplasmic reticulum and mitochondria[45], and plays role in mitochondrial endoxidation via mitochondrial complex I and IV enzyme activity in skin samples[46], which is aligned with *TRH*'s known involvement in hair and skin development (as indicated in our IPA results, Table S7). Variants in *FAM86HP* (Family With Sequence Similarity 86 Member H, Pseudogene) have been reported to be associated with BMI-adjusted waist-hip

ratio[47]. In the pathway analysis (Fig 4), we observed that *FAM86HP* has an indirect interaction with *TGM2* (Transglutaminase 2), a stress-response gene[48] involved in mRNA metabolism[49] which has been reported to be upregulated during inflammation[48]. It is interesting to note that *TGM2* was over expressed in individuals receiving chemo & radiation therapy, suggesting its involvement in sensitivity to radiation[50,51]. Both, *TGM2* and *NUPR1* (Fig 4) have been reported to be implicated in inflammation driven primarily via the JAK/Stat and IL-17A signaling pathway[52]. In the pathway we observe, *NUPR1* has direct interaction with *FAM72C/FAM27D* and LOC100133315, both of which were identified from the copy number variation association. *NUPR1* is known to repair double strand DNA breaks [53] and regulate cell cycle progression[54] from damage induced by gamma irradiation[53]. The under expression of *NUPR1* has been reported to result in increase in ROS production thus creating a deficit in mitochondrial membrane potential. This alteration in OXPHOS activity has been associated with ER stress and triggers programmed necrosis in cancer cells[55]. During angiogenesis in cancer cells, *NUPR1* was reported to be upregulated in association with triiodothyronine thyroid hormone receptor[56], which is regulated with TRH - Thyrotropin releasing hormone[57]. It is interesting to note that *NUPR1* has been reported to play different roles before cancer development, during cancer progression and in response to cancer treatment.

Normal tissues receive varying amount of radiation depending on their proximity to the tumor tissue, thus exhibiting a spectrum of toxicity effects. Genome-wide and molecular studies have shown that alterations in the DNA-damage response (DDR) from ATM (ataxia-telangiectasia mutation) mutations influence intraindividual variations to radiation toxicity[58]. The findings in this study observe that genes such as *ATRIP*, *NUPR1*, *TELO2*, and *TRH* have multiple roles in damage detection and response mechanism. The genes and their PPI networks highlight their involvement in cell cycle arrest upon detection of DNA damage, affecting DNA replication and repair. During

course of DDR, ROS from radiation seems to promote mitochondrial-induced apoptosis[59]. Further, a constant but varying amount of inflammatory responses due to tissue injury (normal and tumor) appears to trigger ROS–induced mitochondrial oxidation[60], which exacerbates local inflammation in nearby tissues and propagates this DNA repair-inflammation stress cycle. Our findings report several novel genes which have been observed to be associated with known BRCA1-ATM-RAD50 damage response complex[61] which are activated in response to radiation, thus extending our understanding of these new players and their multifactorial roles associated with proctitis.

Our study has several limitations. The sample size is small, and our findings should be replicated and functionally validated in future studies. Our study concentrated on prostate cancer survivors of Caucasian ethnicity, hence to understand if similar or different genes are associated with proctitis developed during treatment of other cancers, it is imperative that our methods be applied to other races/ethnicities and other cancers. In the CNV regions, we mapped several noncoding regions, including lcRNA and snRNAs that would require fine mapping to further understand their involvement with proctitis. Unfortunately, there are only handful of studies that investigate genetics underlying radiotoxicity phenotypes.

There are also many strengths to our study. Leveraging SNPs for reference transcriptomic data and copy number association, we identified several novel genes associated with proctitis – an inflammation of the rectum resulting from radiation therapy received for prostate cancer. The integration of tissue - specific PPI network aided in understanding the biological interactions between known and reported genes. Analysis of copy number variation identified several genes in the reported regions and their pathways analysis highlighted two primary genes that have distinctive roles before cancer development and in response to cancer treatment.

In conclusion, this investigation highlights genes primarily involved in DNA repair processes and mitochondrial malfunction threaded via inflammation. The field of radiogenomics –work investigating the role of genetics in developing radiation toxicity—calls for investigation of genetic risk that can help inform dose management of radiation treatment and toxicity monitoring during treatment[8]. We anticipate that understanding genetic data from both CNV and SNPs would contribute towards optimization of radiation treatment on an individual basis. Similar studies in the future would play strong role in early clinical interventions or periodic checkups for individuals who have high expression of DNA damage activity and alterations in copy number within specific regions.

ACKNOWLEDGEMENTS

We would like to acknowledge Dr. Ranajit Chakraborty for his guidance and interest during the start of the project, we unfortunately lost him in September 2018. We appreciate the datasets received via authorized access from dbGaP- GenePARE dataset. We would also like to acknowledge the NIH – Neurobiology of Aging T32 grant AG020494 for supporting this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

GAP conceptualized the study design, carried out analysis and drafted the manuscript. NRP supervised the study, contributed to revisions and gave final approval for publication.

Competing interests

The authors have no competing interests to declare.

Data Availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

Content in Supplementary file

Table S1: Significantly Associated Genes With Proctitis In Prostate Tissue

Table S2: Gene Set Enrichment For Genes Identified In The Prostate Tissue Network In Fig.1

Table S3: Significant Genes Associated With Proctitis In Whole Blood Tissue

Table S4: Gene Set Enrichment For Genes Identified In The Whole Blood Tissue Network In Fig.2

Table S5: Significant Copy Number Variation Associated With Proctitis

Table S6: Gene Annotation Of Significant Cnv Regions From Ucsc Browser (Grch37/Hg19)

Table S7: Significant Functions Identified For The Network Constructed In Ipa® (Fig.4) Of Genes In Cnv Regions

Figure S1: Visual Summary Of Methodology

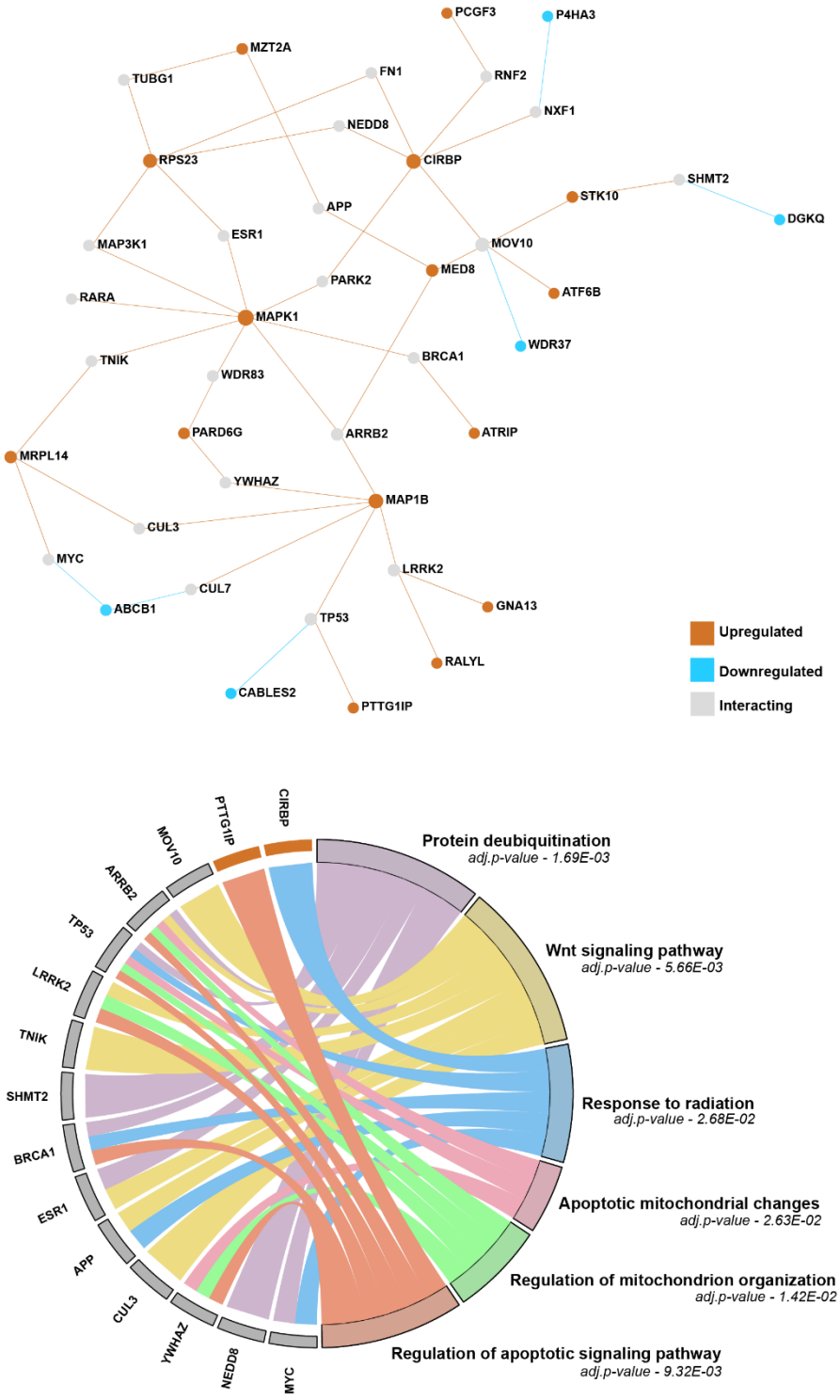


***Figure 1*: Gene Expression of prostate tissue.** *(Top). Genes that are upregulated are shown as orange nodes in the network and downregulated genes are shown as blue nodes. The grey nodes are interacting nodes derived from prostate tissue specific PPI-information. (Bottom). The chord plot summarizes enriched gene ontology pathways of the genes from the network shown in the top panel. The FDR p-value of each pathway is shown under the name of the GO category. See Table S2 in Supplementary file for more details.*

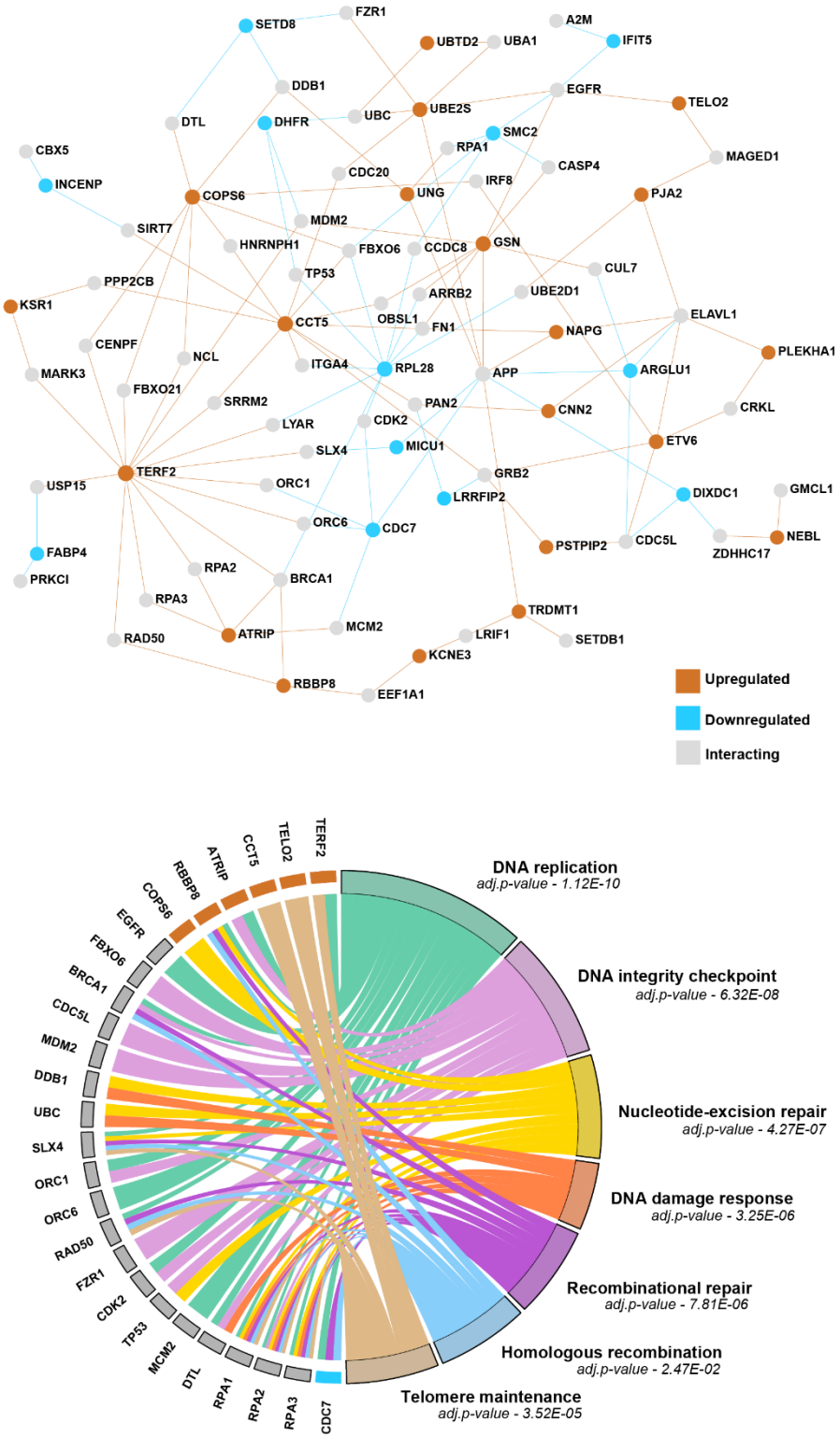


***Figure 211**: **Gene expression for whole blood tissue**. (Top). Genes that are upregulated are shown as orange nodes in the network and downregulated genes are shown as blue nodes. The grey nodes are interacting nodes derived from whole-blood specific PPI-information. (Bottom). The chord plot summarizes enriched gene ontology pathways of the genes from the network shown in the top panel. The FDR p-value of each pathway is shown under the name of the GO category. See Table S4 in Supplementary file for more details.*

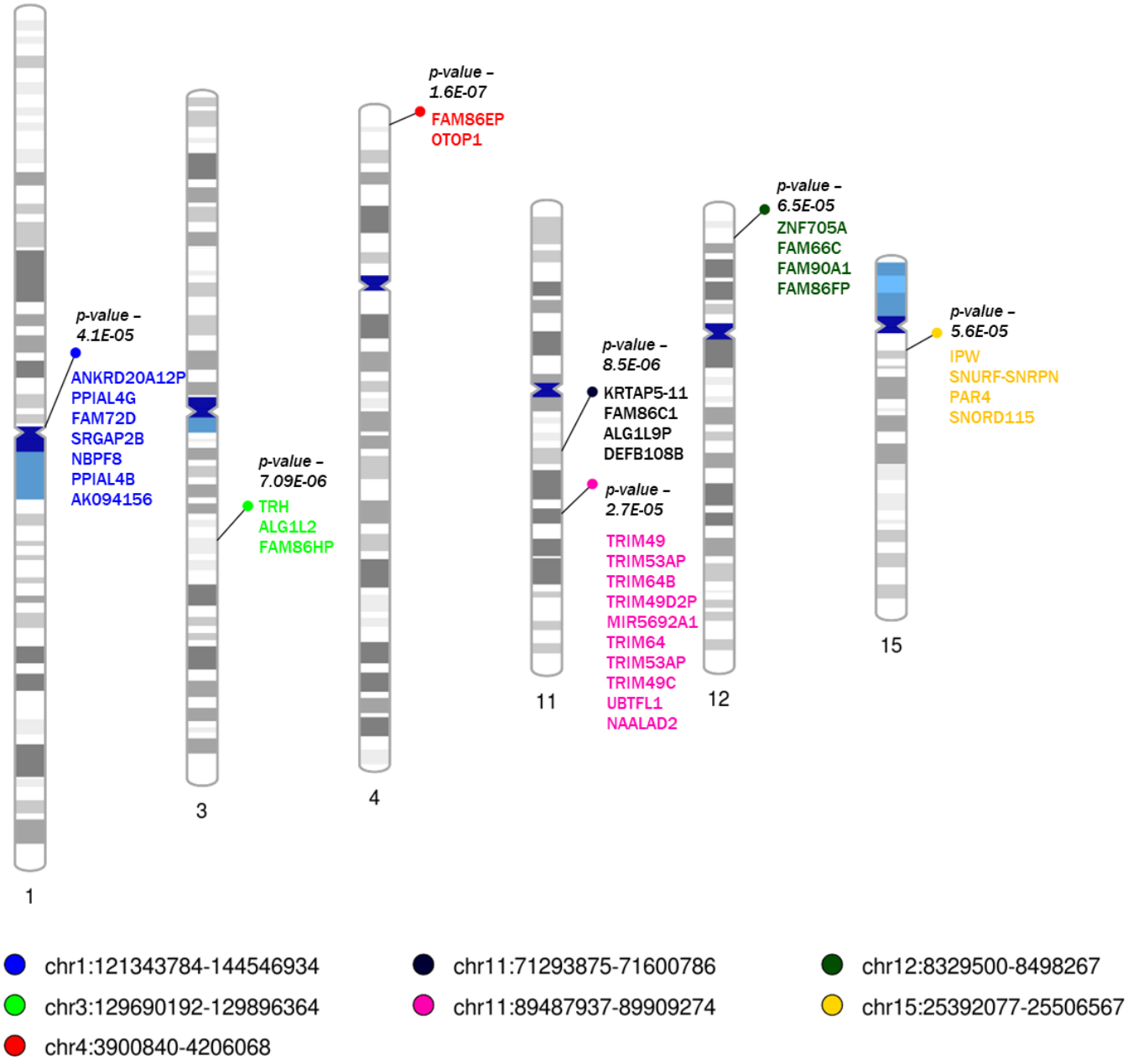


***Figure 3*: Copy Number Variation Analysis.** *Significant CNV regions are labelled on the ideogram with their corresponding p-values. Each region is highlighted with circles and its legend is shown at the bottom. The coding genes within each CNV region is shown in the same color as the regions color from the legend. For full list of annotated molecules, please see Table S6.*

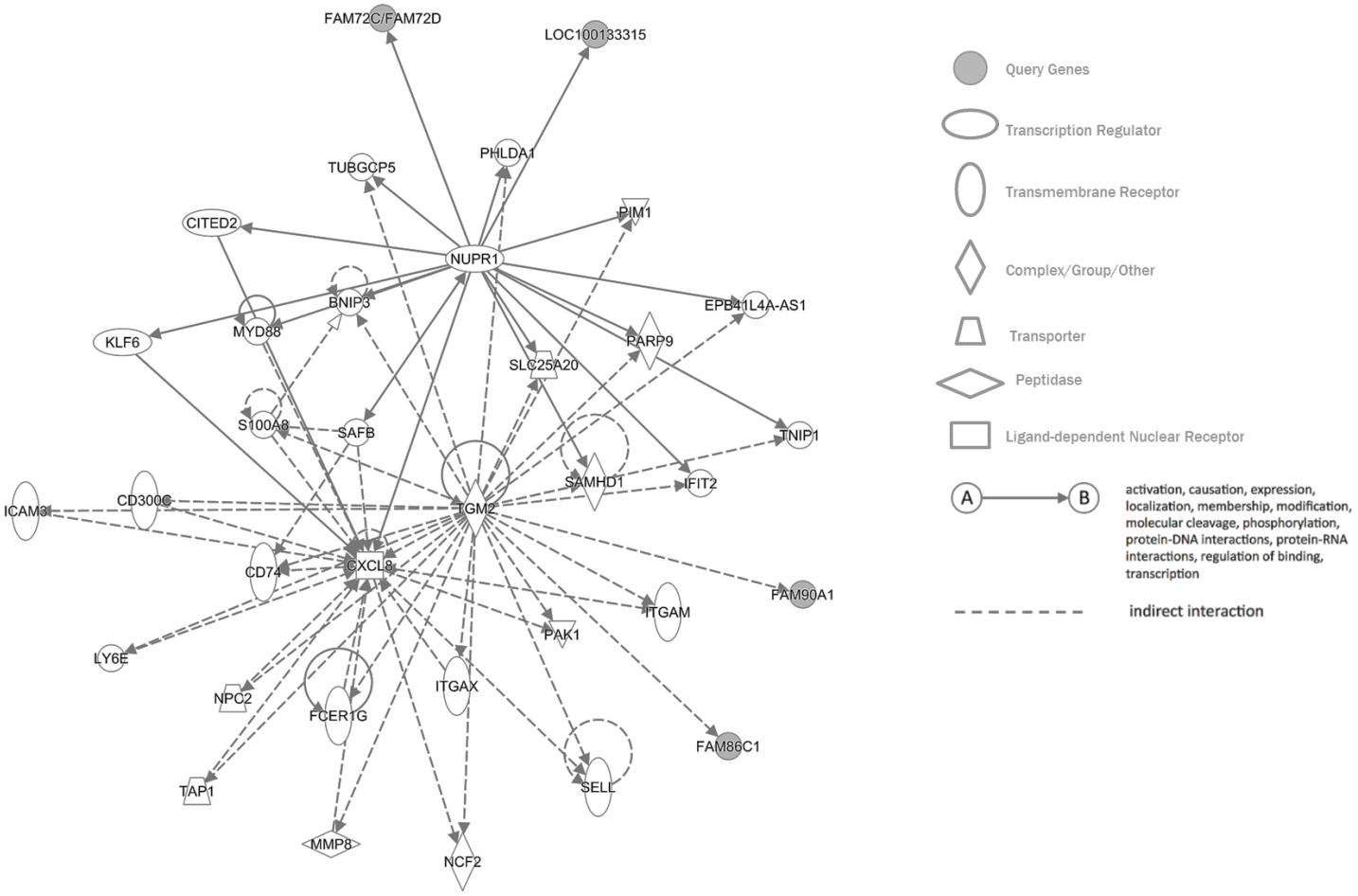


Figure 4: Network identified for genes identified in significant CNV regions. *The network was generated using IPA®, the grey filled circles are query genes from the identified CNV regions. Two genes – NUPR1 and TGM2 were found to interact with the query genes.*

2.5 REFERENCES

1 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2017. **67**, 7-30, doi:10.3322/caac.21387 (2017).

2 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J Clin* **69**, 7-34, doi:10.3322/caac.21551 (2019).

3 Fachal, L. *et al.* A three-stage genome-wide association study identifies a susceptibility locus for late radiotherapy toxicity at 2q24.1. *Nature Genetics* **46**, 891, doi:10.1038/ng.3020

https://www.nature.com/articles/ng.3020#supplementary-information (2014).

4 Do, N. L., Nagle, D. & Poylin, V. Y. Radiation proctitis: current strategies in management. *Gastroenterol Res Pract* **2011**, 917941-917941, doi:10.1155/2011/917941 (2011).

5 Tabaja, L., Sidani, S. M. J. D. D. & Sciences. Management of Radiation Proctitis. **63**, 2180-2188, doi:10.1007/s10620-018-5163-8 (2018).

6 Vanneste, B. G. L. *et al.* Chronic radiation proctitis: tricks to prevent and treat. *Int J Colorectal Dis* **30**, 1293-1303, doi:10.1007/s00384-015-2289-4 (2015).

7 Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* **66**, 271-289, doi:10.3322/caac.21349 (2016).

8 West, C. M. & Barnett, G. C. Genetics and genomics of radiotherapy toxicity: towards prediction. *Genome medicine* **3**, 52, doi:10.1186/gm268 (2011).

9 Zyla, J. *et al.* Combining CDKN1A gene expression and genome-wide SNPs in a twin cohort to gain insight into the heritability of individual radiosensitivity. **19**, 575-585, doi:10.1007/s10142-019-00658-3 (2019).

10 Barnett, G. C. *et al.* A genome wide association study (GWAS) providing evidence of an association between common genetic variants and late radiotherapy toxicity. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* **111**, 178-185, doi:10.1016/j.radonc.2014.02.012 (2014).

11 Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091-1098, doi:10.1038/ng.3367 (2015).

12 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

13 Valsesia, A., Macé, A., Jacquemont, S., Beckmann, J. S. & Kutalik, Z. The Growing Importance of CNVs: New Insights for Detection and Clinical Interpretation. *Frontiers in genetics* **4**, 92-92, doi:10.3389/fgene.2013.00092 (2013).

14 Yang, X. *et al.* Constructing a database for the relations between CNV and human genetic diseases via systematic text mining. **19**, 528, doi:10.1186/s12859-018-2526-2 (2018).

15 Kerns, S. L. *et al.* A 2-stage genome-wide association study to identify single nucleotide polymorphisms associated with development of erectile dysfunction following radiation therapy for prostate cancer. *International journal of radiation oncology, biology, physics* **85**, e21-28, doi:10.1016/j.ijrobp.2012.08.003 (2013).

16 Grodsky, M. B. & Sidani, S. M. Radiation proctopathy. *Clin Colon Rectal Surg* **28**, 103-111, doi:10.1055/s-0035-1547337 (2015).

17 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

18 Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat Protoc* **5**, 1564-1573, doi:10.1038/nprot.2010.116 (2010).

19 Gaspar, H. A., Hübel, C. & Breen, G. Drug Targetor: a web interface to investigate the human druggome for over 500 phenotypes. *Bioinformatics*, doi:10.1093/bioinformatics/bty982 %J Bioinformatics (2018).

20 Basha, O., Shpringer, R., Argov, C. M. & Yeger-Lotem, E. The DifferentialNet database of differential protein-protein interactions in human tissues. *Nucleic Acids Res* **46**, D522-d526, doi:10.1093/nar/gkx981 (2018).

21 Zhou, G. *et al.* NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Research*, doi:10.1093/nar/gkz240 %J Nucleic Acids Research (2019).

22 Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).

23 Walter, W., Sanchez-Cabo, F. & Ricote, M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912-2914, doi:10.1093/bioinformatics/btv300 (2015).

24 Yu, Y. P. *et al.* Genomic Copy Number Variations in the Genomes of Leukocytes Predict Prostate Cancer Clinical Outcomes. *PLOS ONE* **10**, e0135982, doi:10.1371/journal.pone.0135982 (2015).

25 Kim, J. H. *et al.* CNVRuler: a copy number variation-based case-control association analysis tool. *Bioinformatics* **28**, 1790-1792, doi:10.1093/bioinformatics/bts239 (2012).

26 Wolfe, D., Dudek, S., Ritchie, M. D. & Pendergrass, S. A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min* **6**, 18-18, doi:10.1186/1756-0381-6-18 (2013).

27 Kerns, S. L. *et al.* Genome-wide association study identifies a region on chromosome 11q14.3 associated with late rectal bleeding following radiation therapy for prostate cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* **107**, 372-376, doi:10.1016/j.radonc.2013.05.001 (2013).

28 Kerns, S. L. *et al.* Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer. *EBioMedicine* **10**, 150-163, doi:10.1016/j.ebiom.2016.07.022 (2016).

29 Fachin, A. L. *et al.* Gene Expression Profiles in Radiation Workers Occupationally Exposed to Ionizing Radiation. *Journal of Radiation Research* **50**, 61-71, doi:10.1269/jrr.08034 %J Journal of Radiation Research (2009).

30 Logue, M. W. *et al.* An analysis of gene expression in PTSD implicates genes involved in the glucocorticoid receptor pathway and neural responses to stress. *Psychoneuroendocrinology* **57**, 1-13, doi:10.1016/j.psyneuen.2015.03.016 (2015).

31 Zhang, B., Liu, X., Chen, W. & Chen, L. IFIT5 potentiates anti-viral response through enhancing innate immune signaling pathways. *Acta biochimica et biophysica Sinica* **45**, 867-874, doi:10.1093/abbs/gmt088 (2013).

32 Lo, U. G. *et al.* IFNgamma-Induced IFIT5 Promotes Epithelial-to-Mesenchymal Transition in Prostate Cancer via miRNA Processing. *Cancer research* **79**, 1098-1112, doi:10.1158/0008-5472.Can-18-2207 (2019).

33 Komaki, R. *et al.* [Retinal vasculopathy with cerebral leukoencephalopathy carrying TREX1 mutation diagnosed by the intracranial calcification: a case report]. *Rinsho*

shinkeigaku = Clinical neurology **58**, 111-117, doi:10.5692/clinicalneurol.cn-001096 (2018).

34 Shigechi, T. *et al.* ATR-ATRIP kinase complex triggers activation of the Fanconi anemia DNA repair pathway. *Cancer research* **72**, 1149-1156, doi:10.1158/0008-5472.Can-11-2904 (2012).

35 Subramanian, L. & Nakamura, T. M. A kinase-independent role for the Rad3(ATR)-Rad26(ATRIP) complex in recruitment of Tel1(ATM) to telomeres in fission yeast. *PLoS Genet* **6**, e1000839, doi:10.1371/journal.pgen.1000839 (2010).

36 Hilton, B. A. *et al.* ATR Plays a Direct Antiapoptotic Role at Mitochondria, which Is Regulated by Prolyl Isomerase Pin1. *Mol Cell* **60**, 35-46, doi:10.1016/j.molcel.2015.08.008 (2015).

37 Vanpouille-Box, C. *et al.* DNA exonuclease Trex1 regulates radiotherapy-induced tumour immunogenicity. *Nat Commun* **8**, 15618, doi:10.1038/ncomms15618 (2017).

38 Diamond, J. M. *et al.* Exosomes Shuttle TREX1-Sensitive IFN-Stimulatory dsDNA from Irradiated Cancer Cells to DCs. *Cancer immunology research* **6**, 910-920, doi:10.1158/2326-6066.Cir-17-0581 (2018).

39 Garcia-Muse, T. & Boulton, S. J. Distinct modes of ATR activation after replication stress and DNA double-strand breaks in Caenorhabditis elegans. *Embo j* **24**, 4345-4355, doi:10.1038/sj.emboj.7600896 (2005).

40 Nam, S. Y. *et al.* Phosphorylation of CLK2 at serine 34 and threonine 127 by AKT controls cell survival after ionizing radiation. *The Journal of biological chemistry* **285**, 31157-31163, doi:10.1074/jbc.M110.122044 (2010).

41 Brown, M. C. & Gromeier, M. MNK Controls mTORC1:Substrate Association through Regulation of TELO2 Binding with mTORC1. *Cell reports* **18**, 1444-1457, doi:10.1016/j.celrep.2017.01.023 (2017).

42 Quek, H., Lim, Y. C., Lavin, M. F. & Roberts, T. L. PIKKing a way to regulate inflammation. *Immunology and cell biology* **96**, 8-20, doi:10.1111/imcb.1001 (2018).

43 Fernandez, S. V. *et al.* Expression and DNA methylation changes in human breast epithelial cells after bisphenol A exposure. *International journal of oncology* **41**, 369-377, doi:10.3892/ijo.2012.1444 (2012).

44 Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, baw100-baw100, doi:10.1093/database/baw100 (2016).

45 Ronning, S. A., Heatley, G. A. & Martin, T. F. Thyrotropin-releasing hormone mobilizes Ca2+ from endoplasmic reticulum and mitochondria of GH3 pituitary cells: characterization of cellular Ca2+ pools by a method based on digitonin permeabilization. *Proc Natl Acad Sci U S A* **79**, 6294-6298, doi:10.1073/pnas.79.20.6294 (1982).

46 Knuever, J. *et al.* Thyrotropin-releasing hormone controls mitochondrial biology in human epidermis. *J Clin Endocrinol Metab* **97**, 978-986, doi:10.1210/jc.2011-1096 (2012).

47 Southam, L. *et al.* Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. *Nat Commun* **8**, 15606, doi:10.1038/ncomms15606 (2017).

48 Agnihotri, N., Kumar, S. & Mehta, K. Tissue transglutaminase as a central mediator in inflammation-induced progression of breast cancer. *Breast Cancer Res* **15**, 202, doi:10.1186/bcr3371 (2013).

49 Ooko, E., Kadioglu, O., Greten, H. J. & Efferth, T. Pharmacogenomic Characterization and Isobologram Analysis of the Combination of Ascorbic Acid and Curcumin-Two Main Metabolites of Curcuma longa-in Cancer Cells. *Frontiers in pharmacology* **8**, 38, doi:10.3389/fphar.2017.00038 (2017).

50 Leicht, D. T. *et al.* TGM2: a cell surface marker in esophageal adenocarcinomas. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* **9**, 872-881, doi:10.1097/jto.0000000000000229 (2014).

51 Garnier, D. *et al.* Divergent evolution of temozolomide resistance in glioblastoma stem cells is reflected in extracellular vesicles and coupled with radiosensitization. *Neuro-oncology* **20**, 236-248, doi:10.1093/neuonc/nox142 (2018).

52 Rajamani, D. *et al.* Temporal retinal transcriptome and systems biology analysis identifies key pathways and hub genes in Staphylococcus aureus endophthalmitis. *Sci Rep* **6**, 21502, doi:10.1038/srep21502 (2016).

53 Gironella, M. *et al.* p8/nupr1 regulates DNA-repair activity after double-strand gamma irradiation-induced DNA damage. *J Cell Physiol* **221**, 594-602, doi:10.1002/jcp.21889 (2009).

54 Hamidi, T. *et al.* Nupr1-aurora kinase A pathway provides protection against metabolic stress-mediated autophagic-associated cell death. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 5234-5246, doi:10.1158/1078-0432.Ccr-12-0026 (2012).

55 Santofimia-Castano, P. *et al.* Inactivation of NUPR1 promotes cell death by coupling ER-stress responses with necrosis. *Sci Rep* **8**, 16999, doi:10.1038/s41598-018-35020-3 (2018).

56 Chen, C. Y. *et al.* Induction of nuclear protein-1 by thyroid hormone enhances platelet-derived growth factor A mediated angiogenesis in liver cancer. *Theranostics* **9**, 2361-2379, doi:10.7150/thno.29628 (2019).

57 Brent, G. A. Mechanisms of thyroid hormone action. *J Clin Invest* **122**, 3035-3043, doi:10.1172/JCI60047 (2012).

58 De Ruysscher, D. *et al.* Radiotherapy toxicity. *Nature Reviews Disease Primers* **5**, 13, doi:10.1038/s41572-019-0064-5 (2019).

59 Li, X. *et al.* ROS Induced by KillerRed Targeting Mitochondria (mtKR) Enhances Apoptosis Caused by Radiation via Cyt c/Caspase-3 Pathway. *Oxid Med Cell Longev* **2019**, 4528616-4528616, doi:10.1155/2019/4528616 (2019).

60 Zorov, D. B., Juhaszova, M. & Sollott, S. J. Mitochondrial reactive oxygen species (ROS) and ROS-induced ROS release. *Physiol Rev* **94**, 909-950, doi:10.1152/physrev.00026.2013 (2014).

61 Jang, E. R. & Lee, J.-S. DNA damage response mediated through BRCA1. *Cancer research and treatment : official journal of Korean Cancer Association* **36**, 214-221, doi:10.4143/crt.2004.36.4.214 (2004).

2.6 SUPPLEMENTARY FILE

Table S1: *Significantly associated genes with proctitis in prostate tissue*

Gene (Ensemble ID)	Z-score	P-value	Gene Symbol	Gene symbol definition
Downregulated genes based on Z-scores				
ENSG00000149679.7	-3.07	2.15E-03	CABLES2	Cdk5 and Abl enzyme substrate 2 [Source:HGNC Symbol;Acc:16143]
ENSG00000109794.9	-3.04	2.40E-03	FAM149A	family with sequence similarity 149, member A [Source:HGNC Symbol;Acc:24527]
ENSG00000205464.7	-3.01	2.59E-03	ATP6AP1L	ATPase, H+ transporting, lysosomal accessory protein 1-like [Source:HGNC Symbol;Acc:28091]
ENSG00000152778.7	-2.99	2.79E-03	IFIT5	interferon-induced protein with tetratricopeptide repeats 5 [Source:HGNC Symbol;Acc:13328]
ENSG00000270614.1	-2.96	3.03E-03	CTC-325H20.7	NA
ENSG00000235192.1	-2.92	3.50E-03	AC009495.2	NA
ENSG00000230847.4	-2.89	3.79E-03	RP11-195E2.1	NA
ENSG00000225864.1	-2.88	3.97E-03	HCG4P11	HLA complex group 4 pseudogene 11 [Source:HGNC Symbol;Acc:22930]
ENSG00000047056.10	-2.84	4.46E-03	WDR37	WD repeat domain 37 [Source:HGNC Symbol;Acc:31406]
ENSG00000149380.7	-2.79	5.28E-03	P4HA3	prolyl 4-hydroxylase, alpha polypeptide III [Source:HGNC Symbol;Acc:30135]
ENSG00000254974.1	-2.78	5.38E-03	RP11-702H23.2	NA
ENSG00000145214.9	-2.76	5.77E-03	DGKQ	diacylglycerol kinase, theta 110kDa [Source:HGNC Symbol;Acc:2856]
ENSG00000196683.6	-2.75	5.90E-03	TOMM7	translocase of outer mitochondrial membrane 7 homolog (yeast) [Source:HGNC Symbol;Acc:21648]
ENSG00000177144.5	-2.74	6.09E-03	NUDT4P1	nudix (nucleoside diphosphate linked moiety X)-type motif 4 pseudogene 1 [Source:HGNC Symbol;Acc:18012]
ENSG00000234882.1	-2.72	6.50E-03	EIF3EP1	eukaryotic translation initiation factor 3, subunit E pseudogene 1 [Source:HGNC Symbol;Acc:6102]

ENSG00000255987.1	-2.70	7.03E-03	RP11-1094M14.4	NA
ENSG00000136367.12	-2.67	7.66E-03	ZFHX2	zinc finger homeobox 2 [Source:HGNC Symbol;Acc:20152]
ENSG00000167971.14	-2.66	7.76E-03	CASKIN1	CASK interacting protein 1 [Source:HGNC Symbol;Acc:20879]
ENSG00000185482.3	-2.55	1.06E-02	STAC3	SH3 and cysteine rich domain 3 [Source:HGNC Symbol;Acc:28423]
ENSG00000143630.5	-2.52	1.18E-02	HCN3	hyperpolarization activated cyclic nucleotide-gated potassium channel 3 [Source:HGNC Symbol;Acc:19183]
ENSG00000175749.11	-2.47	1.37E-02	EIF3KP1	eukaryotic translation initiation factor 3, subunit K pseudogene 1 [Source:HGNC Symbol;Acc:44016]
ENSG00000188687.11	-2.46	1.38E-02	SLC4A5	solute carrier family 4, sodium bicarbonate cotransporter, member 5 [Source:HGNC Symbol;Acc:18168]
ENSG00000165055.11	-2.45	1.43E-02	METTL2B	methyltransferase like 2B [Source:HGNC Symbol;Acc:18272]
ENSG00000105767.2	-2.43	1.51E-02	CADM4	cell adhesion molecule 4 [Source:HGNC Symbol;Acc:30825]
ENSG00000085563.10	-2.42	1.55E-02	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1 [Source:HGNC Symbol;Acc:40]
ENSG00000174238.10	-2.38	1.73E-02	PITPNA	phosphatidylinositol transfer protein, alpha [Source:HGNC Symbol;Acc:9001]
ENSG00000245556.2	-2.38	1.75E-02	CTD-2037K23.2	NA
ENSG00000101751.6	-2.34	1.95E-02	POLI	polymerase (DNA directed) iota [Source:HGNC Symbol;Acc:9182]
Upregulated genes based on Z-scores				
ENSG00000168993.10	2.33	1.98E-02	CPLX1	complexin 1 [Source:HGNC Symbol;Acc:2309]
ENSG00000256682.1	2.34	1.95E-02	TAS2R12	NA
ENSG00000205809.5	2.34	1.93E-02	KLRC2	killer cell lectin-like receptor subfamily C, member 2 [Source:HGNC Symbol;Acc:6375]
ENSG00000206127.6	2.35	1.90E-02	GOLGA8O	golgin A8 family, member O [Source:HGNC Symbol;Acc:44406]
ENSG00000204520.8	2.36	1.83E-02	MICA	MHC class I polypeptide-related sequence A [Source:HGNC Symbol;Acc:7090]
ENSG00000064886.9	2.37	1.76E-02	CHI3L2	chitinase 3-like 2 [Source:HGNC Symbol;Acc:1933]
ENSG00000178184.11	2.40	1.66E-02	PARD6G	par-6 partitioning defective 6 homolog gamma (C. elegans) [Source:HGNC Symbol;Acc:16076]
ENSG00000259865.1	2.41	1.61E-02	RP11-488L18.10	NA
ENSG00000069493.10	2.42	1.55E-02	CLEC2D	C-type lectin domain family 2, member D [Source:HGNC Symbol;Acc:14351]

ENSG00000100439.6	2.44	1.48E-02	ABHD4	abhydrolase domain containing 4 [Source:HGNC Symbol;Acc:20154]
ENSG00000100030.10	2.45	1.44E-02	MAPK1	mitogen-activated protein kinase 1 [Source:HGNC Symbol;Acc:6871]
ENSG00000104228.8	2.49	1.28E-02	TRIM35	tripartite motif containing 35 [Source:HGNC Symbol;Acc:16285]
ENSG00000260911.1	2.49	1.28E-02	RP11-196G11.2	NA
ENSG00000173272.9	2.50	1.25E-02	MZT2A	mitotic spindle organizing protein 2A [Source:HGNC Symbol;Acc:33187]
ENSG00000185619.13	2.52	1.17E-02	PCGF3	polycomb group ring finger 3 [Source:HGNC Symbol;Acc:10066]
ENSG00000183255.7	2.53	1.14E-02	PTTG1IP	pituitary tumor-transforming 1 interacting protein [Source:HGNC Symbol;Acc:13524]
ENSG00000214376.5	2.55	1.08E-02	VSTM5	V-set and transmembrane domain containing 5 [Source:HGNC Symbol;Acc:34443]
ENSG00000184672.7	2.55	1.07E-02	RALYL	RALY RNA binding protein-like [Source:HGNC Symbol;Acc:27036]
ENSG00000213676.6	2.57	1.02E-02	ATF6B	activating transcription factor 6 beta [Source:HGNC Symbol;Acc:2349]
ENSG00000164053.13	2.60	9.46E-03	ATRIP	ATR interacting protein [Source:HGNC Symbol;Acc:33499]
ENSG00000240494.2	2.62	8.73E-03	RPS12P28	ribosomal protein S12 pseudogene 28 [Source:HGNC Symbol;Acc:36972]
ENSG00000145495.10	2.64	8.39E-03	MARCH6	membrane-associated ring finger (C3HC4) 6, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:30550]
ENSG00000131711.10	2.68	7.31E-03	MAP1B	microtubule-associated protein 1B [Source:HGNC Symbol;Acc:6836]
ENSG00000113430.5	2.70	6.93E-03	IRX4	iroquois homeobox 4 [Source:HGNC Symbol;Acc:6129]
ENSG00000159479.12	2.71	6.77E-03	MED8	mediator complex subunit 8 [Source:HGNC Symbol;Acc:19971]
ENSG00000123545.5	2.71	6.67E-03	NDUFAF4	NADH dehydrogenase (ubiquinone) complex I, assembly factor 4 [Source:HGNC Symbol;Acc:21034]
ENSG00000180992.5	2.84	4.57E-03	MRPL14	mitochondrial ribosomal protein L14 [Source:HGNC Symbol;Acc:14279]
ENSG00000186468.8	2.87	4.07E-03	RPS23	ribosomal protein S23 [Source:HGNC Symbol;Acc:10410]
ENSG00000099622.9	2.89	3.91E-03	CIRBP	cold inducible RNA binding protein [Source:HGNC Symbol;Acc:1982]
ENSG00000244753.2	2.89	3.79E-03	RPL15P21	ribosomal protein L15 pseudogene 21 [Source:HGNC Symbol;Acc:36190]
ENSG00000120063.5	3.04	2.39E-03	GNA13	guanine nucleotide binding protein (G protein), alpha 13 [Source:HGNC Symbol;Acc:4381]
ENSG00000163491.12	3.20	1.39E-03	NEK10	NIMA-related kinase 10 [Source:HGNC Symbol;Acc:18592]

ENSG00000100726.10	3.42	6.29E-04	TELO2	TEL2, telomere maintenance 2, homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:29099]
ENSG00000072786.8	3.44	5.76E-04	STK10	serine/threonine kinase 10 [Source:HGNC Symbol;Acc:11388]

Table S2: *Gene set enrichment for genes identified in the prostate tissue network in Fig.1*

GO ID	Gene Ontology Term	Genes	Adjusted p.val
GO:0016579	protein deubiquitination	ARRB2,TP53,SHMT2,BRCA1,ESR1,NEDD8,MYC	1.69E-03
GO:0016055	Wnt signaling pathway	MOV10,ARRB2,LRRK2,TNIK,ESR1,APP,CUL3	5.66E-03
GO:2001233	regulation of apoptotic signaling pathway	ARRB2,TP53,LRRK2,BRCA1,YWHAZ,PTTG1IP	9.32E-03
GO:0010821	regulation of mitochondrion organization	ARRB2,TP53,LRRK2,YWHAZ	1.42E-02
GO:0008637	apoptotic mitochondrial changes	ARRB2,TP53,YWHAZ	2.63E-02
GO:0009314	response to radiation	CIRBP,TP53,BRCA1,APP,MYC	2.68E-02

Table S3: *Significant genes associated with proctitis in whole blood tissue*

Gene (Ensemble ID)	Z-score	P-value	Gene Symbol	Gene symbol definition
Downregulated genes based on Z-scores				
ENSG00000205464.7	-3.62	2.93E-04	ATP6AP1L	ATPase, H+ transporting, lysosomal accessory protein 1-like [Source:HGNC Symbol;Acc:28091]
ENSG00000106077.14	-3.55	3.92E-04	ABHD11	abhydrolase domain containing 11 [Source:HGNC Symbol;Acc:16407]
ENSG00000116786.7	-3.33	8.63E-04	PLEKHM2	pleckstrin homology domain containing, family M (with RUN domain) member 2 [Source:HGNC Symbol;Acc:29131]
ENSG00000197566.5	-3.13	1.74E-03	ZNF624	zinc finger protein 624 [Source:HGNC Symbol;Acc:29254]
ENSG00000008513.10	-3.11	1.88E-03	ST3GAL1	ST3 beta-galactoside alpha-2,3-sialyltransferase 1 [Source:HGNC Symbol;Acc:10862]
ENSG00000171766.11	-3.11	1.89E-03	GATM	glycine amidinotransferase (L-arginine:glycine amidinotransferase) [Source:HGNC Symbol;Acc:4175]
ENSG00000149679.7	-3.09	1.98E-03	CABLES2	Cdk5 and Abl enzyme substrate 2 [Source:HGNC Symbol;Acc:16143]
ENSG00000137819.9	-2.95	3.17E-03	PAQR5	progestin and adipoQ receptor family member V [Source:HGNC Symbol;Acc:29645]
ENSG00000251287.4	-2.94	3.25E-03	ALG1L2	ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase-like 2 [Source:HGNC Symbol;Acc:37258]
ENSG00000107798.13	-2.89	3.79E-03	LIPA	lipase A, lysosomal acid, cholesterol esterase [Source:HGNC Symbol;Acc:6617]
ENSG00000166669.9	-2.89	3.91E-03	ATF7IP2	activating transcription factor 7 interacting protein 2 [Source:HGNC Symbol;Acc:20397]
ENSG00000197774.8	-2.84	4.44E-03	EME2	essential meiotic endonuclease 1 homolog 2 (S. pombe) [Source:HGNC Symbol;Acc:27289]
ENSG00000097046.8	-2.83	4.66E-03	CDC7	cell division cycle 7 [Source:HGNC Symbol;Acc:1745]
ENSG00000143643.8	-2.81	5.01E-03	TTC13	tetratricopeptide repeat domain 13 [Source:HGNC Symbol;Acc:26204]
ENSG00000100360.10	-2.80	5.06E-03	IFT27	intraflagellar transport 27 homolog (Chlamydomonas) [Source:HGNC Symbol;Acc:18626]

ENSG00000118690.8	-2.73	6.27E-03	ARMC2	armadillo repeat containing 2 [Source:HGNC Symbol;Acc:23045]
ENSG00000211451.7	-2.73	6.37E-03	GNRHR2	gonadotropin-releasing hormone (type 2) receptor 2 [Source:HGNC Symbol;Acc:16341]
ENSG00000106638.11	-2.72	6.47E-03	TBL2	transducin (beta)-like 2 [Source:HGNC Symbol;Acc:11586]
ENSG00000093167.13	-2.71	6.64E-03	LRRFIP2	leucine rich repeat (in FLII) interacting protein 2 [Source:HGNC Symbol;Acc:6703]
ENSG00000104093.9	-2.70	6.83E-03	DMXL2	Dmx-like 2 [Source:HGNC Symbol;Acc:2938]
ENSG00000236056.1	-2.67	7.47E-03	GAPDHP14	glyceraldehyde-3-phosphate dehydrogenase pseudogene 14 [Source:HGNC Symbol;Acc:4160]
ENSG00000166664.9	-2.67	7.59E-03	CHRFAM7A	CHRNA7 (cholinergic receptor, nicotinic, alpha 7, exons 5-10) and FAM7A (family with sequence similarity 7A, exons A-E) fusion [Source:HGNC Symbol;Acc:15781]
ENSG00000170458.9	-2.67	7.69E-03	CD14	CD14 molecule [Source:HGNC Symbol;Acc:1628]
ENSG00000122034.8	-2.65	8.08E-03	GTF3A	general transcription factor IIIA [Source:HGNC Symbol;Acc:4662]
ENSG00000167220.7	-2.57	1.01E-02	HDHD2	haloacid dehalogenase-like hydrolase domain containing 2 [Source:HGNC Symbol;Acc:25364]
ENSG00000142046.10	-2.55	1.08E-02	TMEM91	transmembrane protein 91 [Source:HGNC Symbol;Acc:32393]
ENSG00000178184.11	-2.51	1.20E-02	PARD6G	par-6 partitioning defective 6 homolog gamma (C. elegans) [Source:HGNC Symbol;Acc:16076]
ENSG00000152778.7	-2.50	1.24E-02	IFIT5	interferon-induced protein with tetratricopeptide repeats 5 [Source:HGNC Symbol;Acc:13328]
ENSG00000107745.12	-2.50	1.24E-02	MICU1	mitochondrial calcium uptake 1 [Source:HGNC Symbol;Acc:1530]
ENSG00000215302.4	-2.50	1.25E-02	CTD-3092A11.1	NA
ENSG00000205810.4	-2.48	1.31E-02	KLRC3	killer cell lectin-like receptor subfamily C, member 3 [Source:HGNC Symbol;Acc:6376]
ENSG00000150764.9	-2.46	1.38E-02	DIXDC1	DIX domain containing 1 [Source:HGNC Symbol;Acc:23695]
ENSG00000149503.8	-2.46	1.38E-02	INCENP	inner centromere protein antigens 135/155kDa [Source:HGNC Symbol;Acc:6058]

ENSG00000183955.8	-2.46	1.40E-02	SETD8	SET domain containing (lysine methyltransferase) 8 [Source:HGNC Symbol;Acc:29489]
ENSG00000196189.8	-2.45	1.41E-02	SEMA4A	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4A [Source:HGNC Symbol;Acc:10729]
ENSG00000163684.7	-2.45	1.43E-02	RPP14	ribonuclease P/MRP 14kDa subunit [Source:HGNC Symbol;Acc:30327]
ENSG00000134884.9	-2.44	1.45E-02	ARGLU1	arginine and glutamate rich 1 [Source:HGNC Symbol;Acc:25482]
ENSG00000217930.3	-2.44	1.45E-02	PAM16	presequence translocase-associated motor 16 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:29679]
ENSG00000111325.12	-2.43	1.49E-02	OGFOD2	2-oxoglutarate and iron-dependent oxygenase domain containing 2 [Source:HGNC Symbol;Acc:25823]
ENSG00000108107.8	-2.42	1.54E-02	RPL28	ribosomal protein L28 [Source:HGNC Symbol;Acc:10330]
ENSG00000170323.4	-2.42	1.54E-02	FABP4	fatty acid binding protein 4, adipocyte [Source:HGNC Symbol;Acc:3559]
ENSG00000228716.2	-2.41	1.58E-02	DHFR	dihydrofolate reductase [Source:HGNC Symbol;Acc:2861]
ENSG00000227775.3	-2.40	1.63E-02	RP1-283E3.4	NA
ENSG00000114541.10	-2.40	1.65E-02	FRMD4B	FERM domain containing 4B [Source:HGNC Symbol;Acc:24886]
ENSG00000136824.14	-2.39	1.68E-02	SMC2	structural maintenance of chromosomes 2 [Source:HGNC Symbol;Acc:14011]
ENSG00000000457.9	-2.37	1.76E-02	SCYL3	SCY1-like 3 (S. cerevisiae) [Source:HGNC Symbol;Acc:19285]
ENSG00000242588.2	-2.35	1.87E-02	RP11-274B21.1	NA
ENSG00000141086.13	-2.35	1.88E-02	CTRL	chymotrypsin-like [Source:HGNC Symbol;Acc:2524]
ENSG00000127561.10	-2.35	1.89E-02	SYNGR3	synaptogyrin 3 [Source:HGNC Symbol;Acc:11501]
Upregulated genes based on Z-scores				
ENSG00000204516.5	2.33	1.99E-02	MICB	MHC class I polypeptide-related sequence B [Source:HGNC Symbol;Acc:7091]
ENSG00000239736.2	2.33	1.99E-02	CEACAMP3	carcinoembryonic antigen-related cell adhesion molecule pseudogene 3 [Source:HGNC Symbol;Acc:1825]

ENSG00000175538.6	2.35	1.90E-02	KCNE3	potassium voltage-gated channel, Isk-related family, member 3 [Source:HGNC Symbol;Acc:6243]
ENSG00000155542.7	2.35	1.89E-02	SETD9	SET domain containing 9 [Source:HGNC Symbol;Acc:28508]
ENSG00000104731.9	2.35	1.86E-02	KLHDC4	kelch domain containing 4 [Source:HGNC Symbol;Acc:25272]
ENSG00000244733.4	2.36	1.85E-02	RP11-506M13.3	NA
ENSG00000251584.1	2.36	1.85E-02	RP11-440I14.3	NA
ENSG00000198961.5	2.36	1.84E-02	PJA2	praja ring finger 2, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:17481]
ENSG00000233297.4	2.36	1.82E-02	RASA4DP	RAS p21 protein activator 4CD, pseudogene [Source:HGNC Symbol;Acc:44226]
ENSG00000184350.8	2.37	1.77E-02	MRGPRE	MAS-related GPR, member E [Source:HGNC Symbol;Acc:30694]
ENSG00000173559.8	2.37	1.76E-02	NABP1	nucleic acid binding protein 1 [Source:HGNC Symbol;Acc:26232]
ENSG00000136250.7	2.38	1.74E-02	AOAH	acyloxyacyl hydrolase (neutrophil) [Source:HGNC Symbol;Acc:548]
ENSG00000078114.14	2.39	1.70E-02	NEBL	nebulette [Source:HGNC Symbol;Acc:16932]
ENSG00000076248.6	2.40	1.66E-02	UNG	uracil-DNA glycosylase [Source:HGNC Symbol;Acc:12572]
ENSG00000272787.1	2.41	1.58E-02	NA	NA
ENSG00000154511.7	2.43	1.52E-02	FAM69A	family with sequence similarity 69, member A [Source:HGNC Symbol;Acc:32213]
ENSG00000186283.9	2.43	1.51E-02	TOR3A	torsin family 3, member A [Source:HGNC Symbol;Acc:11997]
ENSG00000260807.2	2.44	1.46E-02	RP11-161M6.2	NA
ENSG00000184752.8	2.46	1.39E-02	NDUFA12	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 12 [Source:HGNC Symbol;Acc:23987]
ENSG00000151470.8	2.47	1.36E-02	C4orf33	chromosome 4 open reading frame 33 [Source:HGNC Symbol;Acc:27025]
ENSG00000143036.12	2.48	1.30E-02	SLC44A3	solute carrier family 44, member 3 [Source:HGNC Symbol;Acc:28689]
ENSG00000139083.6	2.49	1.27E-02	ETV6	ets variant 6 [Source:HGNC Symbol;Acc:3495]

ENSG00000064666.10	2.50	1.26E-02	CNN2	calponin 2 [Source:HGNC Symbol;Acc:2156]
ENSG00000168090.5	2.50	1.23E-02	COPS6	COP9 signalosome subunit 6 [Source:HGNC Symbol;Acc:21749]
ENSG00000101773.12	2.53	1.14E-02	RBBP8	retinoblastoma binding protein 8 [Source:HGNC Symbol;Acc:9891]
ENSG00000107614.17	2.55	1.09E-02	TRDMT1	tRNA aspartic acid methyltransferase 1 [Source:HGNC Symbol;Acc:2977]
ENSG00000263006.2	2.55	1.08E-02	ROCK1P1	Rho-associated, coiled-coil containing protein kinase 1 pseudogene 1 [Source:HGNC Symbol;Acc:37832]
ENSG00000183506.12	2.55	1.06E-02	PI4KAP2	phosphatidylinositol 4-kinase, catalytic, alpha pseudogene 2 [Source:HGNC Symbol;Acc:33577]
ENSG00000117593.8	2.55	1.06E-02	DARS2	aspartyl-tRNA synthetase 2, mitochondrial [Source:HGNC Symbol;Acc:25538]
ENSG00000108106.9	2.58	9.84E-03	UBE2S	ubiquitin-conjugating enzyme E2S [Source:HGNC Symbol;Acc:17895]
ENSG00000168246.5	2.65	7.98E-03	UBTD2	ubiquitin domain containing 2 [Source:HGNC Symbol;Acc:24463]
ENSG00000110274.10	2.67	7.67E-03	CEP164	centrosomal protein 164kDa [Source:HGNC Symbol;Acc:29182]
ENSG00000164053.13	2.68	7.28E-03	ATRIP	ATR interacting protein [Source:HGNC Symbol;Acc:33499]
ENSG00000150753.7	2.71	6.70E-03	CCT5	chaperonin containing TCP1, subunit 5 (epsilon) [Source:HGNC Symbol;Acc:1618]
ENSG00000141068.9	2.74	6.06E-03	KSR1	kinase suppressor of ras 1 [Source:HGNC Symbol;Acc:6465]
ENSG00000100726.10	2.76	5.86E-03	TELO2	TEL2, telomere maintenance 2, homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:29099]
ENSG00000138468.11	2.83	4.63E-03	SENP7	SUMO1/sentrin specific peptidase 7 [Source:HGNC Symbol;Acc:30402]
ENSG00000168016.9	2.85	4.38E-03	TRANK1	tetratricopeptide repeat and ankyrin repeat containing 1 [Source:HGNC Symbol;Acc:29011]
ENSG00000107679.10	2.89	3.87E-03	PLEKHA1	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1 [Source:HGNC Symbol;Acc:14335]
ENSG00000152229.14	2.92	3.50E-03	PSTPIP2	proline-serine-threonine phosphatase interacting protein 2 [Source:HGNC Symbol;Acc:9581]

ENSG00000215908.5	2.94	3.29E-03	CROCCP2	ciliary rootlet coiled-coil, rootletin pseudogene 2 [Source:HGNC Symbol;Acc:28170]
ENSG00000105707.9	3.05	2.32E-03	HPN	hepsin [Source:HGNC Symbol;Acc:5155]
ENSG00000134265.8	3.09	2.02E-03	NAPG	N-ethylmaleimide-sensitive factor attachment protein, gamma [Source:HGNC Symbol;Acc:7642]
ENSG00000148180.12	3.29	9.99E-04	GSN	gelsolin [Source:HGNC Symbol;Acc:4620]
ENSG00000095917.9	3.30	9.66E-04	TPSD1	tryptase delta 1 [Source:HGNC Symbol;Acc:14118]
ENSG00000132604.6	3.42	6.22E-04	TERF2	telomeric repeat binding factor 2 [Source:HGNC Symbol;Acc:11729]
ENSG00000089351.10	3.45	5.61E-04	GRAMD1A	GRAM domain containing 1A [Source:HGNC Symbol;Acc:29305]
ENSG00000105711.6	3.57	3.51E-04	SCN1B	sodium channel, voltage-gated, type I, beta subunit [Source:HGNC Symbol;Acc:10586]
ENSG00000145423.4	3.78	1.55E-04	SFRP2	secreted frizzled-related protein 2 [Source:HGNC Symbol;Acc:10777]

Table S4: *Gene set enrichment for genes identified in the whole blood tissue network in Fig.2*

GO ID	Gene Ontology Term	Genes	Adjusted p.val
GO:0006260	DNA replication	TERF2, EGFR, CDC7, BRCA1, ATRIP, RBBP8, SLX4, ORC1,O RC6,RAD50,CDK2,MCM2,DTL,RPA1,RPA2,RPA3	1.12E-10
GO:0031570	DNA integrity checkpoint	FBXO6, BRCA1, CDC5L, ATRIP, MDM2,ORC1,FZR1,CDK2,TP53,DTL,RPA2	6.32E-08
GO:0006289	nucleotide-excision repair	COPS6,RBBP8,DDB1,UBC,SLX4,TP53,RPA1,RPA2,RPA3	4.27E-07
GO:0042769	DNA damage response, detection of DNA damage	DDB1, UBC,DTL,RPA1,RPA2,RPA3	3.25E-06
GO:0000725	recombinational repair	CDC7,BRCA1,RBBP8,SLX4,RAD50,RPA1,RPA2,RPA3	7.81E-06
GO:0000723	telomere maintenance	TERF2,CCT5,SLX4,RAD50,TELO2,RPA1,RPA2,RPA3	3.52E-05
GO:0035825	homologous recombination	SLX4,RAD50,RPA2	2.47E-02

Table S5: *Significant copy number variation associated with proctitis*

CNVR ID	Chr	Start	End	Size	Description	z value	p value	Odds Ratio	Lower CI	Upper CI	FDR
CNVR_2810_3	4	3900840	4206068	305229	mixed	5.239295	1.61E-07	7.045454545	3.393982117	14.6254	0.000721
CNVR_2535_2	3	129690192	129896364	206173	mixed	4.491091	7.09E-06	4.908854156	2.451449101	9.829635	0.012584
CNVR_7565_1	11	71293875	71600786	306912	mixed	4.453558	8.45E-06	5.115384615	2.494035738	10.4918	0.012584
CNVR_7646_1	11	89487937	89909274	421338	mixed	4.19543	2.72E-05	4.984126984	2.353384475	10.5556	0.030436
CNVR_505_1	1	121343784	144546934	2320315 1	mixed	4.103306	4.07E-05	4.210526316	2.118915294	8.366796	0.036412
CNVR_9133_2	15	25392077	25506567	114491	mixed	4.027933	5.62E-05	4.0625	2.053777422	8.035879	0.041921
CNVR_7863_1	12	8329500	8498267	168768	mixed	3.990438	6.59E-05	4.904214511	2.245843909	10.7092 6	0.042115

Table S6: *Gene annotation of significant CNV regions from UCSC Browser (GRCh37/hg19)*

Chromosome	Strand	Transcription Start Position	Transcription Stop Position	Gene Symbol	Gene Symbol Description
chr4	+	3941972	3941998	DQ584669	Homo sapiens piRNA piR-51781, complete sequence.
chr4	-	3943668	3957148	FAM86EP	Homo sapiens family with sequence similarity 86, member E, pseudogene (FAM86EP), non-coding RNA.
chr4	-	3943668	3957148	FAM86EP	Homo sapiens family with sequence similarity 86, member E, pseudogene (FAM86EP), non-coding RNA.
chr4	+	4034896	4076783	BC042823	Homo sapiens cDNA clone IMAGE:5275587.
chr4	-	4190529	4228621	OTOP1	Homo sapiens otopetrin 1 (OTOP1), mRNA.
chr3	+	129693235	129696781	TRH	Homo sapiens thyrotropin-releasing hormone (TRH), mRNA.
chr3	+	129800673	129817233	ALG1L2	Homo sapiens ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase-like 2 (ALG1L2), mRNA.
chr3	+	129800673	129817233	ALG1L2	Homo sapiens ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase-like 2 (ALG1L2), mRNA.
chr3	-	129816624	129822720	FAM86HP	Homo sapiens family with sequence similarity 86, member H, pseudogene (FAM86HP), non-coding RNA.
chr3	-	129816624	129830276	FAM86HP	Homo sapiens family with sequence similarity 86, member H, pseudogene (FAM86HP), non-coding RNA.
chr11	-	71292900	71293921	KRTAP5-11	Homo sapiens keratin associated protein 5-11 (KRTAP5-11), mRNA.
chr11	+	71498556	71512280	FAM86C1	Homo sapiens family with sequence similarity 86, member C1 (FAM86C1), transcript variant 1, mRNA.
chr11	+	71498556	71512280	FAM86C1	Homo sapiens family with sequence similarity 86, member C1 (FAM86C1), transcript variant 3, mRNA.
chr11	+	71498556	71512280	FAM86C1	Homo sapiens family with sequence similarity 86, member C1 (FAM86C1), transcript variant 2, mRNA.
chr11	+	71498556	71512280	FAM86C1	Homo sapiens family with sequence similarity 86, member C1 (FAM86C1), transcript variant 3, mRNA.
chr11	+	71498556	71512280	FAM86C1	Homo sapiens family with sequence similarity 86, member C1 (FAM86C1), transcript variant 2, mRNA.

chr11	-	71505408	71524905	ALG1L9P	Homo sapiens asparagine-linked glycosylation 1-like 9, pseudogene (ALG1L9P), transcript variant 3, non-coding RNA.
chr11	-	71506860	71524905	ALG1L9P	Homo sapiens asparagine-linked glycosylation 1-like 9, pseudogene (ALG1L9P), transcript variant 1, non-coding RNA.
chr11	-	71521023	71524905	ALG1L9P	Homo sapiens asparagine-linked glycosylation 1-like 9, pseudogene (ALG1L9P), transcript variant 2, non-coding RNA.
chr11	+	71544245	71548608	DEFB108B	Homo sapiens defensin, beta 108B (DEFB108B), mRNA.
chr11	-	71576554	71639493	LOC100133315	Homo sapiens transient receptor potential cation channel, subfamily C, member 2-like (LOC100133315), non-coding RNA.
chr11	+	71589498	71595607	LOC100129216	Homo sapiens beta-defensin 131-like (LOC100129216), mRNA.
chr11	-	89530822	89541743	TRIM49	Homo sapiens tripartite motif containing 49 (TRIM49), mRNA.
chr11	+	89553492	89559667	LOC642414	Homo sapiens clone E1 LOC642414 pseudogene mRNA, partial sequence.
chr11	+	89575164	89584136	TRIM53AP	Homo sapiens tripartite motif containing 53A, pseudogene (TRIM53AP), non-coding RNA.
chr11	-	89603605	89609185	TRIM64B	Homo sapiens tripartite motif containing 64B (TRIM64B), mRNA.
chr11	-	89644578	89653576	TRIM49D2P	Homo sapiens tripartite motif containing 49D2, pseudogene (TRIM49D2P), mRNA.
chr11	+	89657231	89666229	TRIM49D2P	Homo sapiens tripartite motif containing 49D2, pseudogene (TRIM49D2P), mRNA.
chr11	+	89673687	89673722	MIR5692A1	Homo sapiens microRNA 5692a-1 (MIR5692A1), microRNA.
chr11	+	89701671	89707240	TRIM64	Homo sapiens tripartite motif containing 64 (TRIM64), mRNA.
chr11	-	89726708	89735676	TRIM53AP	Homo sapiens tripartite motif containing 53A, pseudogene (TRIM53AP), non-coding RNA.
chr11	-	89751201	89757434	LOC642414	Homo sapiens clone E1 LOC642414 pseudogene mRNA, partial sequence.
chr11	+	89764273	89775193	TRIM49C	Homo sapiens tripartite motif containing 49C (TRIM49C), mRNA.
chr11	-	89794201	89796159	LOC440061	Homo sapiens unknown mRNA sequence.
chr11	+	89819117	89820299	UBTFL1	Homo sapiens upstream binding transcription factor, RNA polymerase I-like 1 (UBTFL1), mRNA.

chr11	+	89867817	89892505	NAALAD2	Homo sapiens N-acetylated alpha-linked acidic dipeptidase 2 (NAALAD2), mRNA.
chr11	+	89867817	89897369	NAALAD2	Homo sapiens N-acetylated alpha-linked acidic dipeptidase 2 (NAALAD2), mRNA.
chr11	+	89867817	89925779	NAALAD2	Homo sapiens N-acetylated alpha-linked acidic dipeptidase 2 (NAALAD2), mRNA.
chr11	+	89867817	89925779	NAALAD2	Homo sapiens N-acetylated alpha-linked acidic dipeptidase 2 (NAALAD2), mRNA.
chr11	+	89867817	89925779	NAALAD2	Homo sapiens N-acetylated alpha-linked acidic dipeptidase 2 (NAALAD2), mRNA.
chr1	+	142618798	143257763	CR936796	Homo sapiens PNAS-130 mRNA, complete cds.
chr1	+	142660105	142660135	DQ579288	Homo sapiens piRNA piR-47400, complete sequence.
chr1	-	142672190	142672219	DQ586768	Homo sapiens piRNA piR-53880, complete sequence.
chr1	+	142688238	142688268	DQ579288	Homo sapiens piRNA piR-47400, complete sequence.
chr1	-	142689205	142689235	DQ583161	Homo sapiens piRNA piR-50272, complete sequence.
chr1	+	142689523	142689553	DQ590589	Homo sapiens piRNA piR-32372, complete sequence.
chr1	-	142697420	142713605	ANKRD20A12P	Homo sapiens ankyrin repeat domain 20 family, member A12, pseudogene (ANKRD20A12P), non-coding RNA.
chr1	-	142803223	142888797	BC053679	Homo sapiens cDNA: FLJ22715 fis, clone HSI13726.
chr1	+	142803530	142826645	BC071797	Homo sapiens, clone IMAGE:4720764, mRNA.
chr1	-	142804898	142890670	BC029473	Homo sapiens cDNA clone IMAGE:4723680, **** WARNING: chimeric clone ****.
chr1	-	142851394	142851424	DQ590126	Homo sapiens piRNA piR-47400, complete sequence.
chr1	+	142853227	142855999	DQ592442	Homo sapiens cDNA FLJ35140 fis, clone PLACE6009524.
chr1	+	143119060	143163748	AK056396	Homo sapiens PNAS-130 mRNA, complete cds.
chr1	-	143168647	143168675	DQ586768	Homo sapiens piRNA piR-53880, complete sequence.
chr1	+	143184707	143184737	DQ579288	Homo sapiens piRNA piR-47400, complete sequence.
chr1	+	143185980	143186010	DQ590589	Homo sapiens piRNA piR-32372, complete sequence.

chr1	-	143286955	143286985	DQ587539	Homo sapiens piRNA piR-32372, complete sequence.
chr1	-	143289062	143289092	DQ590126	Homo sapiens piRNA piR-47400, complete sequence.
chr1	-	143402283	143402313	DQ587539	Homo sapiens piRNA piR-32372, complete sequence.
chr1	-	143403553	143403583	DQ590126	Homo sapiens piRNA piR-47400, complete sequence.
chr1	+	143419286	143419314	DQ596206	Homo sapiens piRNA piR-53880, complete sequence.
chr1	-	143424332	143467651	BC070106	Homo sapiens cDNA clone IMAGE:30343207.
chr1	-	143431367	143431397	DQ590126	Homo sapiens piRNA piR-47400, complete sequence.
chr1	-	143647638	143744587	LINC00875	Homo sapiens long intergenic non-protein coding RNA 875 (LINC00875), non-coding RNA.
chr1	+	143687129	143705973	LOC100130000	Homo sapiens phosphodiesterase 4D interacting protein pseudogene (LOC100130000), non-coding RNA.
chr1	+	143687129	143714180	LOC100130000	Homo sapiens phosphodiesterase 4D interacting protein pseudogene (LOC100130000), non-coding RNA.
chr1	+	143690027	143690101	TRNA_Asn	transfer RNA Asn (anticodon GTT)
chr1	+	143702303	143702331	DQ571491	Homo sapiens piRNA piR-31603, complete sequence.
chr1	-	143717587	143744587	LINC00875	Homo sapiens long intergenic non-protein coding RNA 875 (LINC00875), non-coding RNA.
chr1	-	143718512	143744587	LINC00875	Homo sapiens long intergenic non-protein coding RNA 875 (LINC00875), non-coding RNA.
chr1	-	143719238	143744587	LINC00875	Homo sapiens long intergenic non-protein coding RNA 875 (LINC00875), non-coding RNA.
chr1	-	143767143	143767881	PPIAL4G	Homo sapiens peptidylprolyl isomerase A (cyclophilin A)-like 4G (PPIAL4G), mRNA.
chr1	-	143879831	143879905	TRNA_Asn	transfer RNA Asn (anticodon GTT)
chr1	-	143896451	143913143	FAM72D	Homo sapiens family with sequence similarity 72, member D (FAM72D), mRNA.
chr1	-	143896451	143913143	FAM72D	Homo sapiens family with sequence similarity 72, member D (FAM72D), mRNA.
chr1	+	143915747	144094477	SRGAP2B	Homo sapiens SLIT-ROBO Rho GTPase activating protein 2B (SRGAP2B), mRNA.

chr1	+	144146810	144830407	NBPF8	Homo sapiens neuroblastoma breakpoint family, member 8 (NBPF8), transcript variant 3, non-coding RNA.
chr1	+	144146810	146467744	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144146810	144830407	NBPF9	Homo sapiens neuroblastoma breakpoint family, member 9 (NBPF9), transcript variant 2, mRNA.
chr1	+	144148789	144830407	NBPF8	Homo sapiens neuroblastoma breakpoint family, member 8 (NBPF8), transcript variant 3, non-coding RNA.
chr1	+	144150981	144167711	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144151518	144830407	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144160410	144201052	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144162000	144167711	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144162000	144186823	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144162887	144182059	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144171303	144172451	AF420437	Homo sapiens AB13 precursor RNA, partial cds.
chr1	+	144176438	144830407	NBPF8	Homo sapiens neuroblastoma breakpoint family, member 8 (NBPF8), transcript variant 3, non-coding RNA.
chr1	+	144179473	144223374	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144181950	144186823	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144183587	144830407	NBPF8	Homo sapiens neuroblastoma breakpoint family, member 8 (NBPF8), transcript variant 3, non-coding RNA.
chr1	+	144184251	145339512	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144185827	144223374	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.

chr1	+	144189011	144827928	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144190581	144830407	NBPF8	Homo sapiens neuroblastoma breakpoint family, member 8 (NBPF8), transcript variant 3, non-coding RNA.
chr1	+	144190581	144209025	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144196230	144201052	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144196230	144209025	LOC100288142	Homo sapiens neuroblastoma breakpoint family member (LOC100288142), mRNA.
chr1	+	144209457	144212180	AL050141	Homo sapiens mRNA; cDNA DKFZp586O031 (from clone DKFZp586O031).
chr1	+	144218498	144830407	NBPF8	Homo sapiens neuroblastoma breakpoint family, member 8 (NBPF8), transcript variant 3, non-coding RNA.
chr1	-	144275887	144290006	AX746564	Homo sapiens cDNA FLJ33341 fis, clone BRACE2002582.
chr1	-	144300511	144340773	LINC00623	Homo sapiens long intergenic non-protein coding RNA 623 (LINC00623), non-coding RNA.
chr1	-	144300511	144340773	LINC00623	Homo sapiens long intergenic non-protein coding RNA 623 (LINC00623), non-coding RNA.
chr1	-	144300511	144341755	LOC728875	Homo sapiens uncharacterized LOC728875 (LOC728875), non-coding RNA.
chr1	+	144301610	144301684	TRNA_Asn	transfer RNA Asn (anticodon GTT)
chr1	-	144308613	144308687	TRNA_Asn	transfer RNA Asn (anticodon GTT)
chr1	-	144339562	144340773	LINC00623	Homo sapiens long intergenic non-protein coding RNA 623 (LINC00623), non-coding RNA.
chr1	-	144340773	144341077	BC047032	Homo sapiens cDNA FLJ33341 fis, clone BRACE2002582.
chr1	-	144363461	144364246	PPIAL4B	Homo sapiens peptidylprolyl isomerase A (cyclophilin A)-like 4B (PPIAL4B), mRNA.
chr1	-	144456162	144470244	AX746564	Homo sapiens cDNA FLJ33341 fis, clone BRACE2002582.
chr1	-	144480744	144521009	LOC728875	Homo sapiens uncharacterized LOC728875 (LOC728875), non-coding RNA.

chr1	-	144480745	144521969	LOC728875	Homo sapiens uncharacterized LOC728875 (LOC728875), non-coding RNA.
chr1	-	144480745	144521009	LOC728875	Homo sapiens uncharacterized LOC728875 (LOC728875), non-coding RNA.
chr1	+	144481839	144481913	TRNA_Asn	transfer RNA Asn (anticodon GTT)
chr1	-	144488842	144488916	TRNA_Asn	transfer RNA Asn (anticodon GTT)
chr1	-	144514872	144519720	AK094156	Homo sapiens primary neuroblastoma cDNA, clone:Nbla04072, full insert sequence.
chr15	+	25362556	25420017	IPW	Homo sapiens imprinted in Prader-Willi syndrome (non-protein coding) (IPW), non-coding RNA.
chr15	+	25415869	25415951	SNORD115-1	Homo sapiens small nucleolar RNA, C/D box 115-1 (SNORD115-1), small nucleolar RNA.
chr15	+	25417781	25417863	SNORD115-2	Homo sapiens small nucleolar RNA, C/D box 115-2 (SNORD115-2), small nucleolar RNA.
chr15	+	25418125	25427252	SNURF-SNRPN	Homo sapiens clone Rt-5 SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25420073	25420155	SNORD115-3	Homo sapiens small nucleolar RNA, C/D box 115-3 (SNORD115-3), small nucleolar RNA.
chr15	+	25421978	25422060	SNORD115-4	Homo sapiens small nucleolar RNA, C/D box 115-4 (SNORD115-4), small nucleolar RNA.
chr15	+	25423884	25423966	SNORD115-5	Homo sapiens small nucleolar RNA, C/D box 115-5 (SNORD115-5), small nucleolar RNA.
chr15	+	25425643	25425725	SNORD115-6	Homo sapiens small nucleolar RNA, C/D box 115-6 (SNORD115-6), small nucleolar RNA.
chr15	+	25426956	25430721	SNURF-SNRPN	Homo sapiens clone Rt-7 SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25427531	25427613	SNORD115-7	Homo sapiens small nucleolar RNA, C/D box 115-7 (SNORD115-7), small nucleolar RNA.
chr15	+	25427858	25431154	SNURF-SNRPN	Homo sapiens clone Rt-9 SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25429452	25429534	SNORD115-8	Homo sapiens small nucleolar RNA, C/D box 115-8 (SNORD115-8), small nucleolar RNA.

chr15	+	25430777	25430859	SNORD115-9	Homo sapiens small nucleolar RNA, C/D box 115-9 (SNORD115-9), small nucleolar RNA.
chr15	+	25432682	25432763	SNORD115-10	Homo sapiens small nucleolar RNA, C/D box 115-10 (SNORD115-10), small nucleolar RNA.
chr15	+	25434560	25434642	SNORD115-11	Homo sapiens small nucleolar RNA, C/D box 115-11 (SNORD115-11), small nucleolar RNA.
chr15	+	25436386	25442649	SNURF-SNRPN	Homo sapiens clone Rt-11 SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25436562	25436644	SNORD115-9	Homo sapiens small nucleolar RNA, C/D box 115-9 (SNORD115-9), small nucleolar RNA.
chr15	+	25438467	25438549	SNORD115-13	Homo sapiens small nucleolar RNA, C/D box 115-13 (SNORD115-13), small nucleolar RNA.
chr15	+	25440067	25440148	SNORD115-14	Homo sapiens small nucleolar RNA, C/D box 115-14 (SNORD115-14), small nucleolar RNA.
chr15	+	25442574	25459165	SNURF-SNRPN	Homo sapiens clone Rt-13I SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25442574	25459165	SNURF-SNRPN	Homo sapiens clone Rt-13I SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25444594	25444676	SNORD115-16	Homo sapiens small nucleolar RNA, C/D box 115-16 (SNORD115-16), small nucleolar RNA.
chr15	+	25446469	25446551	SNORD115-17	Homo sapiens small nucleolar RNA, C/D box 115-17 (SNORD115-17), small nucleolar RNA.
chr15	+	25448373	25448455	SNORD115-17	Homo sapiens small nucleolar RNA, C/D box 115-17 (SNORD115-17), small nucleolar RNA.
chr15	+	25449503	25449585	SNORD115-17	Homo sapiens small nucleolar RNA, C/D box 115-17 (SNORD115-17), small nucleolar RNA.
chr15	+	25451408	25477615	SNORD115-15	Homo sapiens small nucleolar RNA, C/D box 115-15 (SNORD115-15), small nucleolar RNA.
chr15	+	25455064	25455146	SNORD115-22	Homo sapiens small nucleolar RNA, C/D box 115-22 (SNORD115-22), small nucleolar RNA.
chr15	+	25456838	25457180	PAR4	Homo sapiens Prader-Willi/Angelman region gene 4 (PAR4), non-coding RNA.

chr15	+	25458605	25467437	SNURF-SNRPN	Homo sapiens clone Rt-13I SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25458805	25458876	SNORD115-24	Homo sapiens small nucleolar RNA, C/D box 115-24 (SNORD115-24), small nucleolar RNA.
chr15	+	25460687	25460769	SNORD115-25	Homo sapiens small nucleolar RNA, C/D box 115-25 (SNORD115-25), small nucleolar RNA.
chr15	+	25463498	25479764	SNURF-SNRPN	Homo sapiens clone Rt-15 SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25463763	25463845	SNORD115-26	Homo sapiens small nucleolar RNA, C/D box 115-26 (SNORD115-26), small nucleolar RNA.
chr15	+	25465649	25465725	SNORD115-27	Homo sapiens small nucleolar RNA, C/D box 115-27 (SNORD115-27), small nucleolar RNA.
chr15	+	25467500	25467574	SNORD115-28	Homo sapiens small nucleolar RNA, C/D box 115-28 (SNORD115-28), small nucleolar RNA.
chr15	+	25468392	25468474	SNORD115-11	Homo sapiens small nucleolar RNA, C/D box 115-11 (SNORD115-11), small nucleolar RNA.
chr15	+	25470349	25470431	SNORD115-30	Homo sapiens small nucleolar RNA, C/D box 115-30 (SNORD115-30), small nucleolar RNA.
chr15	+	25472255	25472337	SNORD115-31	Homo sapiens small nucleolar RNA, C/D box 115-31 (SNORD115-31), small nucleolar RNA.
chr15	+	25474113	25474195	SNORD115-32	Homo sapiens small nucleolar RNA, C/D box 115-32 (SNORD115-32), small nucleolar RNA.
chr15	+	25475984	25476066	SNORD115-33	Homo sapiens small nucleolar RNA, C/D box 115-33 (SNORD115-33), small nucleolar RNA.
chr15	+	25479393	25479475	SNORD115-35	Homo sapiens small nucleolar RNA, C/D box 115-35 (SNORD115-35), small nucleolar RNA.
chr15	+	25481231	25481313	SNORD115-11	Homo sapiens small nucleolar RNA, C/D box 115-11 (SNORD115-11), small nucleolar RNA.
chr15	+	25481555	25620623	SNURF-SNRPN	Homo sapiens clone Rt-16 SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced.
chr15	+	25483132	25483214	SNORD115-37	Homo sapiens small nucleolar RNA, C/D box 115-37 (SNORD115-37), small nucleolar RNA.

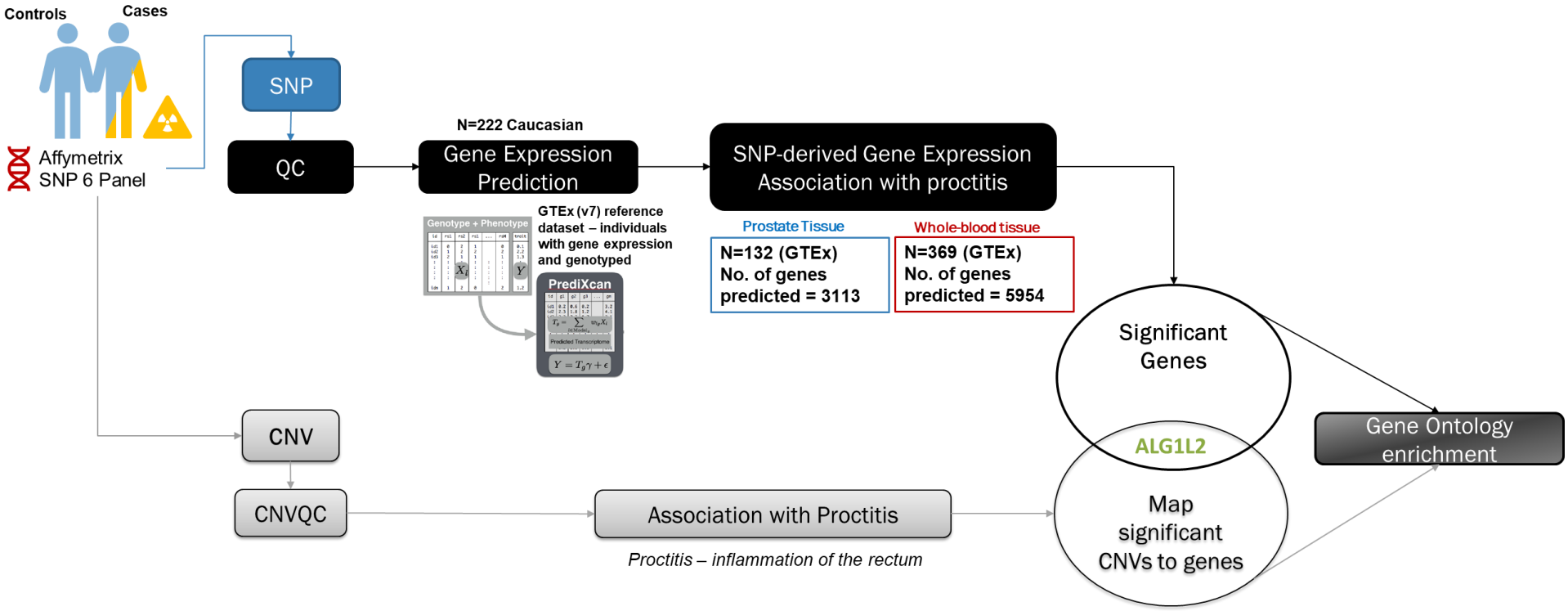
chr15	+	25484984	25485066	SNORD115-38	Homo sapiens small nucleolar RNA, C/D box 115-38 (SNORD115-38), small nucleolar RNA.
chr15	+	25486892	25486974	SNORD115-39	Homo sapiens small nucleolar RNA, C/D box 115-39 (SNORD115-39), small nucleolar RNA.
chr15	+	25488760	25488842	SNORD115-40	Homo sapiens small nucleolar RNA, C/D box 115-40 (SNORD115-40), small nucleolar RNA.
chr15	+	25490624	25490706	SNORD115-41	Homo sapiens small nucleolar RNA, C/D box 115-41 (SNORD115-41), small nucleolar RNA.
chr15	+	25492491	25492573	SNORD115-10	Homo sapiens small nucleolar RNA, C/D box 115-10 (SNORD115-10), small nucleolar RNA.
chr15	+	25494344	25494426	SNORD115-11	Homo sapiens small nucleolar RNA, C/D box 115-11 (SNORD115-11), small nucleolar RNA.
chr15	+	25496005	25496087	SNORD115-44	Homo sapiens small nucleolar RNA, C/D box 115-44 (SNORD115-44), small nucleolar RNA.
chr12	+	8325149	8332642	ZNF705A	Homo sapiens zinc finger protein 705A (ZNF705A), mRNA.
chr12	+	8332804	8353596	FAM66C	Homo sapiens family with sequence similarity 66, member C (FAM66C), non-coding RNA.
chr12	+	8332804	8353596	FAM66C	Homo sapiens family with sequence similarity 66, member C (FAM66C), non-coding RNA.
chr12	+	8332804	8356982	FAM66C	Homo sapiens family with sequence similarity 66, member C (FAM66C), non-coding RNA.
chr12	+	8332804	8368747	FAM66C	Homo sapiens family with sequence similarity 66, member C (FAM66C), non-coding RNA.
chr12	-	8373855	8380214	FAM90A1	Homo sapiens family with sequence similarity 90, member A1 (FAM90A1), mRNA.
chr12	-	8373855	8380214	FAM90A1	Homo sapiens family with sequence similarity 90, member A1 (FAM90A1), mRNA.
chr12	-	8383644	8395542	FAM86FP	Homo sapiens family with sequence similarity 86, member F, pseudogene (FAM86FP), non-coding RNA.

Table S7: *Significant functions identified for the network constructed in IPA® (Fig.4) of genes in CNV regions*

Categories	Functions	Diseases or Functions Annotation	p-Value	Molecules
Hair and Skin Development and Function	anagen	Anagen	6.97E-04	TRH
Endocrine System Disorders, Hereditary Disorder, Metabolic Disease, Organismal Injury and Abnormalities	thyrotropin-releasing hormone deficiency	Thyrotropin-releasing hormone deficiency	6.97E-04	TRH
Molecular Transport	transmembrane transport	Transmembrane transport of H+	1.66E-02	OTOP1
Cell Death and Survival, Organismal Injury and Abnormalities	apoptosis	Apoptosis of keratinocytes	2.00E-02	TRH
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	uterine carcinoma	Uterine carcinoma	2.03E-02	KRTAP5-11, FAM72C/FAM72D, OTOP1, FAM90A1, TRIM49/TRIM49C, DEFB108B, TRIM64/TRIM64B
Cancer, Organismal Injury and Abnormalities	adenoma formation	Adenoma	2.53E-02	TRH, NBPF10 (includes others)
Developmental Disorder, Hereditary Disorder, Organismal Injury and Abnormalities	Prader-Willi syndrome	Prader-Willi syndrome	2.55E-02	IPW
Cancer, Endocrine System Disorders, Organismal Injury and Abnormalities, Reproductive System Disease	pituitary gland adenoma	Pituitary gland adenoma	3.23E-02	TRH
Carbohydrate Metabolism, Small Molecule Biochemistry	synthesis	Synthesis of inositol phosphate	3.43E-02	TRH
Amino Acid Metabolism, Small Molecule Biochemistry	synthesis	Synthesis of amino acids	3.70E-02	NAALAD2
Cell-To-Cell Signaling and Interaction, Embryonic Development	response	Response of embryonic cell lines	3.83E-02	TRIM49/TRIM49C
Cell-To-Cell Signaling and Interaction, Hair and Skin Development and Function	response	Response of epithelial cell lines	3.83E-02	TRIM49/TRIM49C
Cell-To-Cell Signaling and Interaction, Renal and Urological System Development and Function	response	Response of kidney cell lines	3.83E-02	TRIM49/TRIM49C
Cellular Development, Cellular Growth and Proliferation, Connective Tissue Development and	proliferation	Proliferation of keratinocytes	4.30E-02	TRH

Function, Hair and Skin Development and Function, Organ Development, Tissue Development				

Figure S112: *Visual summary of methodology*



Chapter 3

INVESTIGATING HYPERTENSION AS A SOURCE OF VASCULAR DEMENTIA OR A COMORBIDITY TO ALZHEIMER'S DISEASE

ABSTRACT. Age-related comorbidity is common and significantly increases the burden for healthcare of the elderly. Alzheimer's disease (AD) – a type of dementia– and hypertension are the two most prevalent age-related conditions and are highly comorbid. While hypertension is a risk factor for vascular dementia (VaD) – a distinct subtype of dementia– the presence of hypertension with AD (AD+Hyp+) is often characterized as probable vascular dementia. In the absence of imaging and other diagnostic tests, differentiating the two pathological states is difficult.

Our goals are to (1) identify differences in vascular dementia profiles, if any, between individuals who have AD only (AD+Hyp-), and individuals with AD+Hyp+ using CSF levels of amyloid β, tau and p-tau, and (2) compare genome-wide DNA profiles of AD+Hyp- and AD+Hyp+ with an unaffected control population. We hypothesize that genetic variants underlying AD+Hyp+ comorbidity pattern will be different than variants associated with VaD, AD, or hypertension independently.

Genotype and clinical data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) were used to conduct comorbidity analyses comparing healthy controls to AD+/Hyp- vs AD+/Hyp+. We compared the CSF biomarkers in three cohorts using one-way ANOVA. We, then evaluated genome wide profiles in three groups, and mapped SNPs to genes based on position and lowest p-value. The significant genes are further examined for co-expression patterns and known disease networks.

Through this exploratory study using a novel cohort stratification design, we highlight the genetic differences in clinically similar phenotypes, indicating the utility of genetic profiling in aiding differential diagnosis of AD+Hyp+ and VaD.

Short Title: Investigating Alzheimer's- Hypertension comorbidity

3.1 INTRODUCTION

In the recent years, there has been a huge shift in the demographics of the population ages 65 years and older, which is expected to double by 2060 reaching 98 million [1]. Alzheimer's disease (AD) is a neurodegenerative condition affecting 5.5 million people worldwide, with the majority of the affected population being 65 years and older [2]. According to the 2011 Alzheimer's Association report, hypertension is the most prevalent comorbid condition, affecting 60% of the Alzheimer's population [2]. While hypertension is a known risk factor for vascular dementia – a subtype of dementia [3] – the presence of hypertension with AD is often categorized as vascular dementia (VD) [4]. Clinical diagnosis of vascular cognitive decline is associated with presence of cerebral vascular damage from subcortical ischemia resulting in symptoms of small vessel disease including infarcts and white matter hyperintensity (WMH)[5]. These lesions are detected using MRI/tomography images of the brain and signify regions of the brain with water content indicating possible damage to axons and brain tissue. Increase in age is associated with enlarged white matter lesions followed by hypertension[6]. However, studies have shown that presence of hypertension presents different risk ratios in AD and vascular dementia. In a longitudinal study of 8.2 years, the risk of systolic blood pressure in causing dementia was higher for vascular dementia than AD[7]. Hypertension has been associated with increased risk of cognitive impairment in AD[8]. A study investigating multiple cardiovascular factors including hypertension as risk factors to multiple neurodegenerative diseases in ~17,000 individuals reported hypertension as the highest risk to vascular dementia followed by AD[9]. However, not all individuals diagnosed with AD have hypertension. In this study, we hypothesize that hypertension as a comorbidity to AD is different than its association with vascular dementia. We first investigated characteristics that are known vascular dementia risk characteristics prevalent in individuals diagnosed with Alzheimer's disease, with and without

hypertension. Studies have shown that clinical vascular dementia can be diagnosed with the cerebellar spinal fluids (CSF) biomarkers - amyloid β, tau and p-tau, by using the following calculation[10].

$$\text{Vascular Dementia (VD)} = \frac{\text{tau(T)} \cdot \text{tau(p}-181)}{\text{A}\beta 42}$$

We evaluated vascular dementia profiles between individuals with AD only (ADHyp-), and individuals with AD and hypertension (ADHyp+). However, VD profiles in our cohort is indistinguishable between individuals of AD with or without hypertension (Hyp), requiring investigation of other possible pathogenic causes. We, therefore hypothesized that genetic variants underlying ADHyp+ comorbidity pattern, will be different than AD only pathology.

.

3.2 METHODS & MATERIALS

We received authorized access to Alzheimer's Disease Imaging Initiative (ADNI), which contains clinical and genotype data of AD individuals. The research protocol for this project was reviewed by University of North Texas Health Science Center Institutional Review Board on June 24, 2016 and determined to be exempt human subject research under IRB–2016-090.

We compared CSF biomarkers – amyloid β, tau and p-tau in three cohorts using one-way ANOVA. For this pilot study phase, we used ADNI-1[11] (a subset of the data) to conduct a gene-based GWAS comparing genome-wide profiles using 535,762 SNP markers in 677 individuals (after QC)on Illumina – 610W genotyping platform. We compared genome-wide profiles using multinomial regression (Trinculo[12]), to compare multi-category disease populations (ADHyp- and ADHyp+) with a control population (unaffected by either AD or hypertension). The associations were

adjusted for age, sex, and the first two principal components of genetic similarity as covariates. In addition to the overall *p*-value for the likelihood ratio test, we identified individual *p*-value for each disease using Wald statistics. We then used *p*-values to run a gene-based association testing using VEGAS 2.0[13], which maps suggestive SNPs (*p*-value <0.05) to genes-based position (± 50kb). The most significant SNPs annotated to genes was visualized in LocusZoom for a 100kb window. We further investigated these top gene hits using the DisGeNET [14] database for associated diseases and visualized using diseasegent2r, an R package for overlapping diseases in gene-disease associations. We additionally used GeneMania[15] to (1) identify top hits from each association that have been reported to be co-expressed, and (2) to visualize gene networks that have common co-expression. ADHyp- vs ADHyp+ groups were compared using MAGMA [16] which uses a principal component-based regression approach to conduct gene associations. We investigated the neighboring genes mapped to the reported variants as candidates for differences in gene expression in independent cohort. We used the second phase of the ADNI cohort, ADNI-2/GO to assess the mean differences in gene expression between disease status groups. Continuous variables were analyzed using one-way ANOVA followed by Tukey's post hoc for pairwise comparisons. Categorical variables were analyzed using chi-square test, and p-value < 0.05 was considered significant.

3.3 RESULTS

We first compared known clinical characteristics of vascular dementia in AD individuals with normal (ADHyp-) and high blood pressure (ADHyp+) along with Alzheimer's disease. Very few cases of history of stroke and prior neurological symptoms were present in all three groups. However, upon comparing white matter hyperintensity (WMH), we observe significant differences between ADHyp- vs ADHyp+, and control vs ADHyp+ group (Figure 1). The age of disease onset

for Alzheimer's disease in the ADHyp- and ADHyp+ group and age in controls group was not significantly different in the pairwise comparison (Figure 2). We investigated possible correlation trends between a) metabolic variables – white matter hyperintensity, body mass index, triglycerides, cholesterol, and systolic blood pressure, b) cerebral spinal fluid (CSF) biomarkers – Aβ, tau, p-tau, and alpha-synuclein and c) other dementia risk factors – age and MMSE score in ADNI-1 cohort. We observe strong positive correlation between the CSF biomarkers – Aβ, tau and p-tau. All three were weakly negatively correlated with systolic blood pressure. While Aβ was positively (weak) correlated with MMSE scores, tau and p-tau were negatively correlated (weak) with MMSE scores. Although WMH is a sign of cerebrovascular damage caused by stroke or hypertension, we did not observe significant correlation between systolic blood pressure and WMH; we did observe a positive correlation (weak)of WMH with age and negative correlation (weak) with MMSE scores (Figure 3). The vascular dementia profiles between controls, ADHyp- and ADHyp+ group were not statistically different between AD groups, however both groups were different when compared to control individuals (Figure 4). The gene based GWAS identified *KMO, TOMM40,* and *PML* for control vs ADHyp- group, and genes – *TCTE1, PML, GFER, STOML1, TBL3, NOXO1, SYNGR3, NPW, ZNF598, NPW* and *SLC9A3R2* for control vs ADHyp+. The gene-based association *detected UBE4B, TINAG, PRRX2, TUBE1, RGR* and *TMX1* for ADHyp- vs ADHyp+ group. Genes identified in each of comparisons were investigated further to determine existing relationships using GeneMania. We observed that several of the genes were co-expressed and shared protein domains and pathways (Figure 5). Our study found several novel genes for the association of AD and AD-hypertension comorbidity. Using the DisGeNET database, we observe that *PML, KMO* and *TOMM40* are associated with other neuropathological disorders such as depression, and schizophrenia[14] (Figure 6). Furthermore, the genes identified for control

vs AD-hypertension comorbidity were found to be associated with various cognitive decline symptoms, and *SLC9A3R2* was associated with hypertension (Figure 7)[14]. The gene-based GWAS is comprised of SNPs which are annotated to genes within a ±50kb window. In order to test that our identified genes (based on differences in variant frequency between controls and AD-hypertension comorbidity) are potentially different in three groups, we did a candidate gene expression study using a separate set of individuals – ADNI-2/GO cohort. We compared normalized RPM values (Reads per million mapped reads) between controls, AD and AD-hypertension group in the ADNI-2/GO. We first compared the WMH profile in ADNI-2/GO cohort, wherein we observed similar trend of higher WMH in individuals with AD-hypertension comorbidity (Figure 8). We investigated genes which were significantly associated between control and ADHyp+ individuals. *PML* was found to be associated in both pairwise comparisons between control vs ADHyp- and control vs ADHyp+. The most significant SNP for *PML* gene was rs1052242, the gene expression between the disease status was not significant, but after sex-based stratification, significant differences were seen between males with ADHyp+ and females with ADHyp- (Figure 9). Another gene, *KMO* is located on chromosome 1, and the most significant SNP was rs1932441, significant differences in *KMO* gene expression were observed between males with ADHyp+ and females with ADHyp- (Figure 10). The rs8045288 on chromosome 16 was annotated to multiple genes, herein we investigated gene expression of neighboring genes around the SNP. We found significant differences between ADHyp- and ADHyp+ for *GFER, SLC9A3R2, ZNF598* and *NOXO1*. We did not see differences for *NPW, NTHL1, TBL3, and SYNGR3* (Figure 11).

3.4 DISCUSSION

This study identified possible differences between vascular dementia and AD-hypertension comorbidity. We observed that known risk factors of vascular dementia were not significantly present in individuals with AD-hypertension comorbidity except for differences in white matter hyperintensity. While WMH was found to be correlated with the expected dementia risk factors – age and decline in MMSE scores, it was not correlated with systolic pressure. Furthermore, while the CSF biomarkers- amyloid β, tau and p-tau are known for their diagnostic contribution in vascular dementia, their profile is unaltered in Alzheimer's-hypertension comorbidity, providing the motivation for investigating other possible pathogenic causes. Through this exploratory study using a novel cohort stratification design, several suggestive signals emerged, although none of the genes reach genome-wide significance (likely due to the small number of individuals). Our results point to known genes as well as several novel genes which are known to have role in multiple CNS disorders. When comparing the control individuals with AD+Hyp- individuals, we observe *TOMM40, PML* and *KMO* as top hits. Gene-disease associations from DisGeNET database highlight their common association to psychiatric disorders. Interestingly, when comparing Controls vs AD+Hyp+, we observe several genes in the chromosome 16 region. While not all the genes implicated in the signal were present in the DisGeNET database, three of the four genes were associated with CNS disorders and SLC9A3R2 is associated with hypertension. We also conducted data mining to identify gene expression patterns within our top gene set using GeneMANIA; remarkably, all top hits were known to be co-expressed via other intermediate genes. PML was a top hit, in both AD+Hyp- and AD+Hyp+ when compared to controls; although the gene network in these two disease states is distinct, there is commonality of PML gene which is known to be responsible for cancer and various CNS disorders [17]. We also observed that gene

expression of *PML, KMO, GFER* and *SLC9A3R2* are significantly different in the three groups, the stratification based on sex documents indicates potential differences in sub-groups. *PML* was first identified in acute promyelocytic leukemia, thus deriving its gene name. *PML* has been reported to be expressed in brain regions including hippocampus, cortex, cerebellum and neuronal cells of the brain. *PML* dysregulation has been associated with synaptic plasticity, abnormal circadian rhythms and frontotemporal dementia. In the aging neuron, an overexpression of PML has been suspected to provide neuroprotection by clearing accumulation of toxic proteins [17]. *KMO* - kynurenine 3-monooxygenase, is involved in kynurenine pathway for tryptophan degeneration, and is primary source of investigation in neurodegenerative conditions such as Huntington's disease[18]. Dysregulation in the kynurenine pathway leads to decreased tryptophan levels which is inversely correlation with the progression of hypertension in rat models[19].In AD patients, metabolites of kynurenine pathway including tryptophan were lower than age-matched controls[20]. *GFER* – growth factor erv1-like, encodes ALR protein (Augmenter of Liver Regeneration) which is primarily localized in the mitochondria. Mutations in this gene has been suspected to be influenced by histone acetylation and reported to cause infantile mitochondrial disorder. The over expression of ALR is reported to lower mitochondrial ROS damage, reduce apoptosis, fibrosis, ER stress and inflammation[21]. *SLC9A3R2* - Solute Carrier Family 9 Subfamily A , encodes NHERF2 gene responsible for increased sodium-hydrogen exchange in rat model of hypertension[22]. Furthermore, variants in *SLC9A3R2* have been associated with systolic and diastolic pressure in GWAS of 750,000 individuals[23].The NOXO1- NADPH oxidase organizer 1 is located on the cellular membrane and its phosphorylation by protein kinase A leads to superoxide anion production by NOX1 and NOX3. The over expression of NOXO1 has been suspected to be associated with high ROS production in the activated microglia[24].

The molecular function of the identified genes indicates presence of oxidative stress and neurodegeneration. The association of SLC9A3R2 in ADHyp+ is consistent with its previously known association with hypertension. Additionally, we observe an overlapping theme of reported genes to regulate ROS damage. Our findings indicate possible involvement of dysregulated oxidation and tryptophan associated with hypertension, furthermore contributing to neurodegeneration.

There are several limitations to this study: our sample size was small and increasing the number of subjects would provide power to detect other genes responsible for differences in AD and AD-hypertension status. Differences in gene expression could be attributed to methylation changes, therefore comparing epigenomic profiles cross-sectionally or longitudinally would provide deeper insights into the comorbid pathogenicity. While studies have reported that hypertension is a stronger risk to vascular dementia than AD, our findings indicate that known risk factors of vascular dementia are less prevalent (or absent) in our cohort of AD-hypertension comorbidity. This observation along with different significant genes in ADHyp- and ADHyp+ profiles leads us to postulate that AD-hypertension comorbidity is different than vascular dementia. Our results did not support relationship between systolic blood pressure, WMH and MMSE, indicating that there could be other factors associated with WMH. One study found association of diastolic and not systolic blood pressure with WMH[6]. Therefore, it is imperative that we conduct future studies to identify other factors that may be playing a role to exacerbate WMHs, and how the reported genes contribute towards phenotypic heterogeneity of AD and AD-hypertension disease type. We believe this study not only enhances risk stratification for AD-hypertension comorbidity but also sheds light on the underlying pathogenesis.

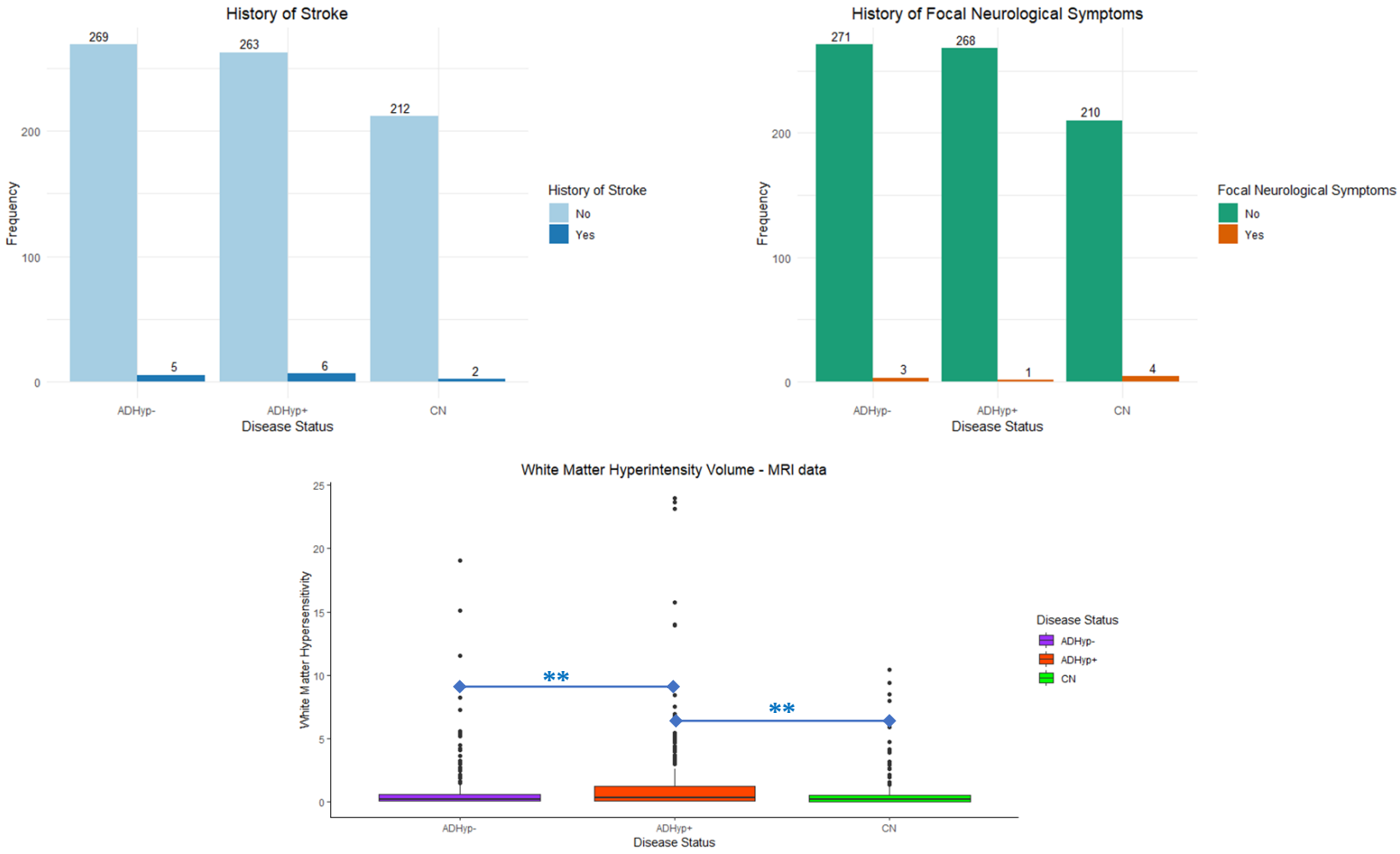


Figure 13: *Characteristics in ADNI-1 population for history of stroke, neurological symptoms and white hyperintensity. The traits were compared between controls (CN), Alzheimer's disease without hypertension (ADHyp-) and Alzheimer's disease with hypertension comorbidity (ADHyp+). Tukey's post-hoc revealed significant differences between ADHyp- vs ADHyp+ (p = 0.005), and CN vs ADHyp+ group (p=0.001).*

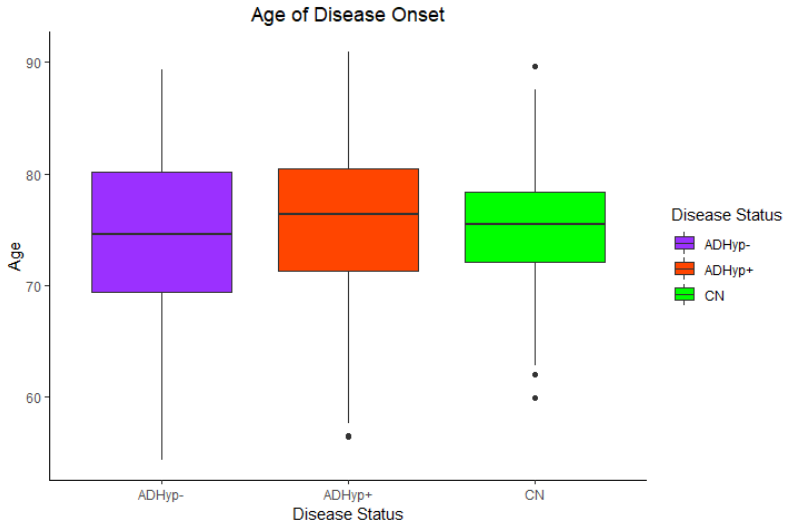


Figure 15: *The age of Alzheimer's disease onset compared within controls (CN), Alzheimer's disease without hypertension (ADHyp-) and Alzheimer's disease with hypertension comorbidity (ADHyp+) in ADNI-1 cohort. The mean differences in age between the groups is not statistically significant.*

CN vs ADHyp+ (p=0.99); ADHyp- vs ADHyp+ (p=0.06); CN vs ADHyp- (p=0.07)

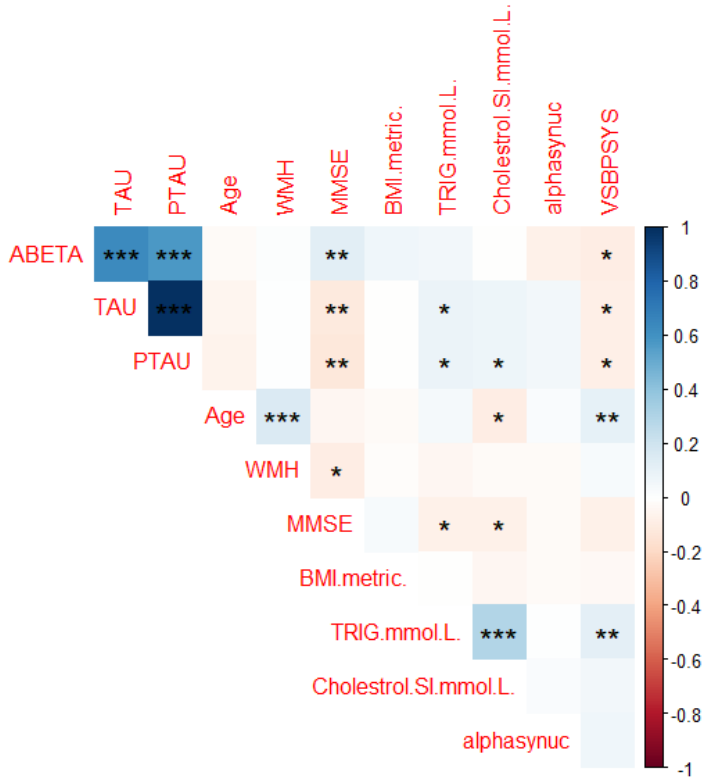


Figure 14: *Correlation plot between metabolic variables – white matter hyperintensity (WMH), body mass index (BMI), triglycerides (TRIG), cholesterol, and systolic blood pressure (VSBPSYS), cerebral spinal fluid (CSF) biomarkers – Aβ, tau, p-tau, and alpha-synuclein and risk factors – age and MMSE score in ADNI-1 cohort. The color scale signifies correlation coefficient, and asterisk represent significance for p-values <0.05 *; <0.01 **; <0.001 ****

$$\text{Vascular Dementia (VD)} = \frac{\text{tau(T)} \cdot \text{tau(p}-181)}{\text{A}\beta 42}$$

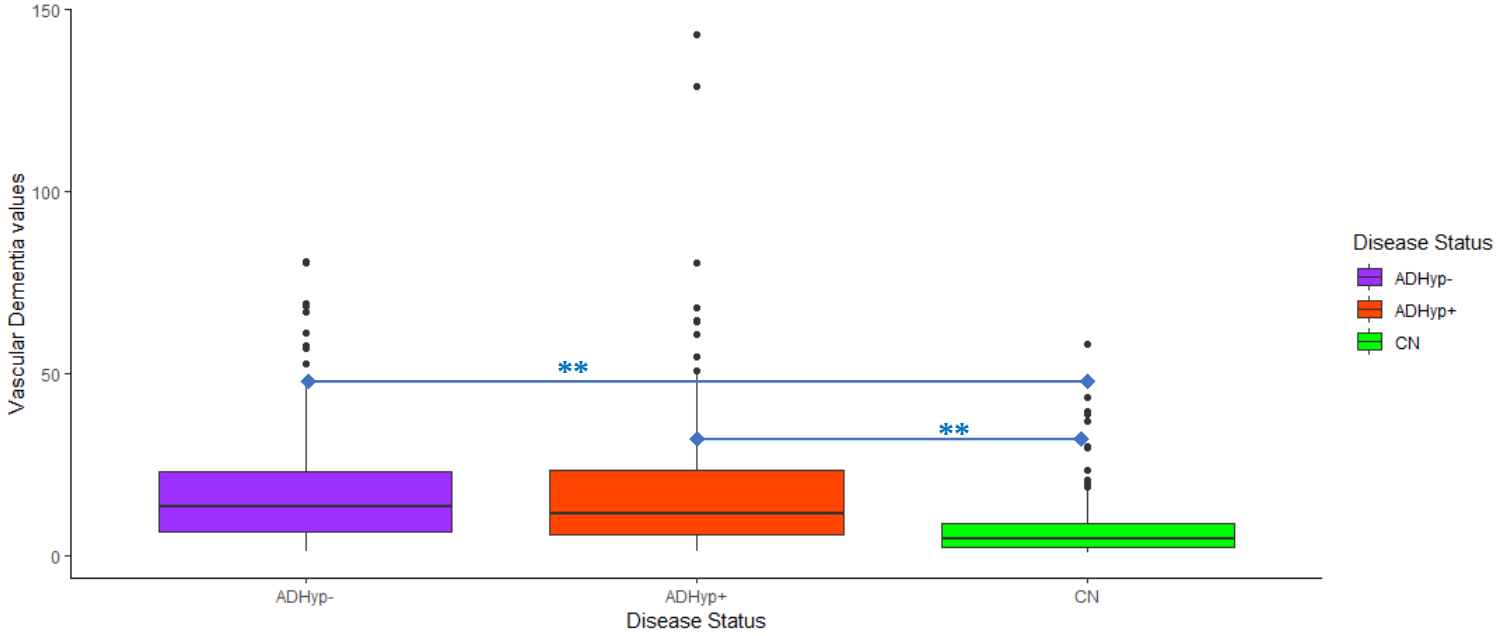


Figure 16: *The vascular dementia profile based on CSF biomarkers – Aβ, tau and p-tau, in controls, Alzheimer's disease without hypertension (Alz) and Alzheimer's disease with hypertension comorbidity (AlzHyp). The pairwise comparison between the three groups reveals control vs AlzHyp and control vs Alz is significant ($p<0.05$) and Alz vs Alzhyp is not significant ($p>0.05$).*

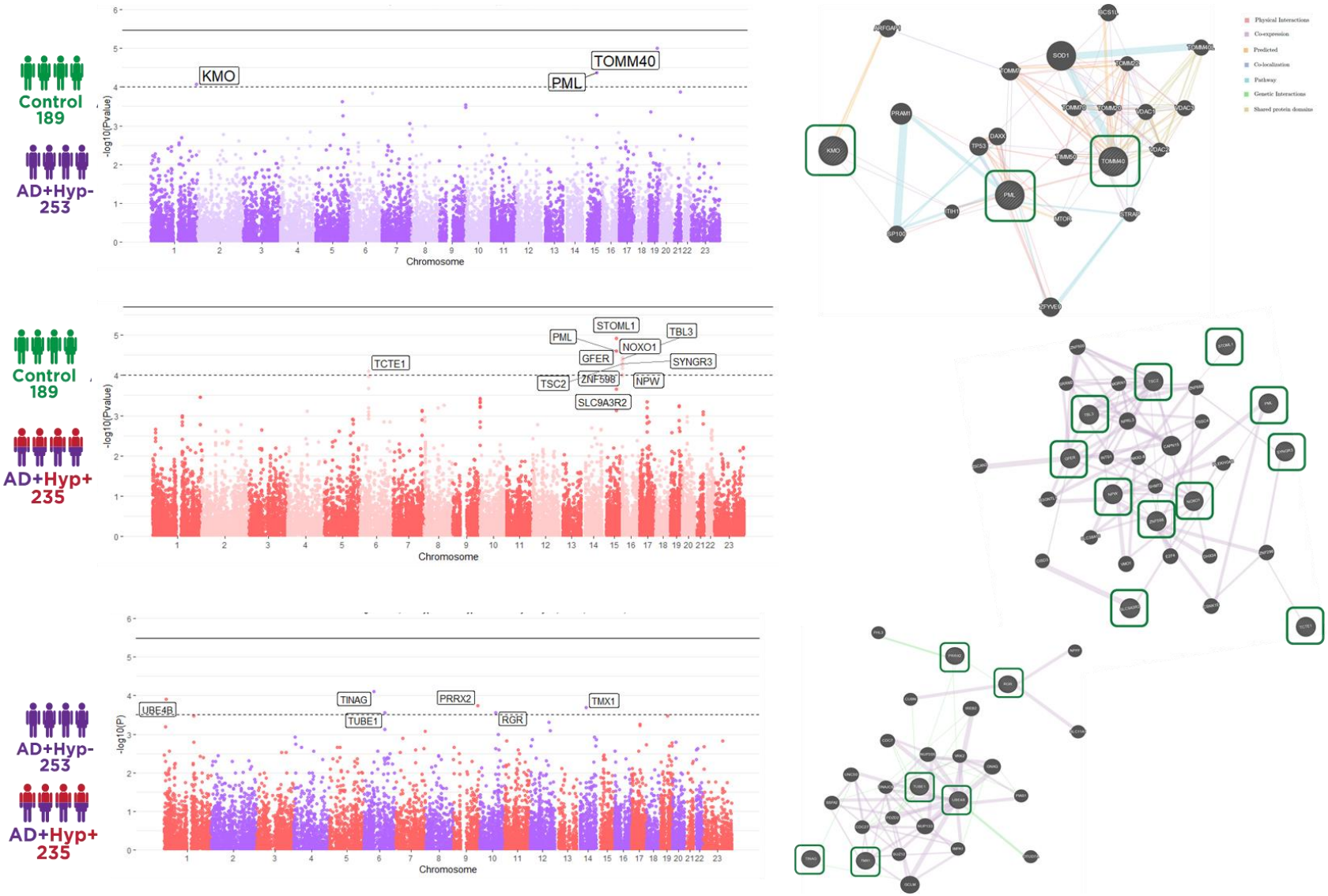


Figure 17: *Gene-based GWAS. Manhattan plot for gene-based association between control vs Alzheimer's disease without hypertension (top), control vs Alzheimer's disease with hypertension comorbidity (middle), and Alzheimer's disease without hypertension vs Alzheimer's disease with hypertension comorbidity (bottom). The genes that are suggestively significant are labelled with the names. Network relationship of the labelled genes are shown next to the corresponding Manhattan plot. The query genes are highlighted in green squares. The log10(p-value) is shown on the y-axis; the dashed lines represent suggestive significance and solid lines represent genome-wide significance based on Bonferroni correction.*

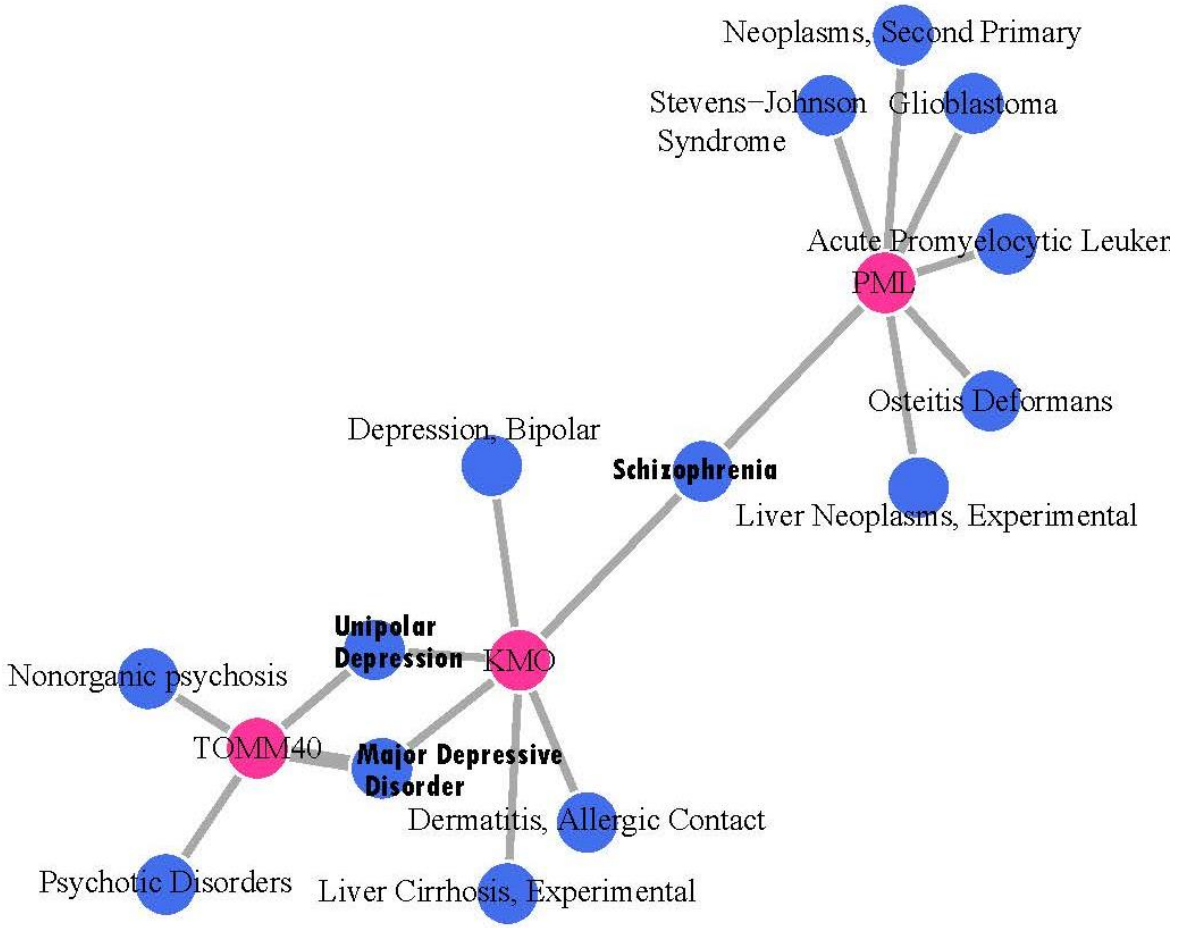


Figure 18: *Disease-gene network for genes identified for the association between control vs Alzheimer's disease without hypertension. The diseases (blue circles) overlapping between genes (pink circles) are bolded.*

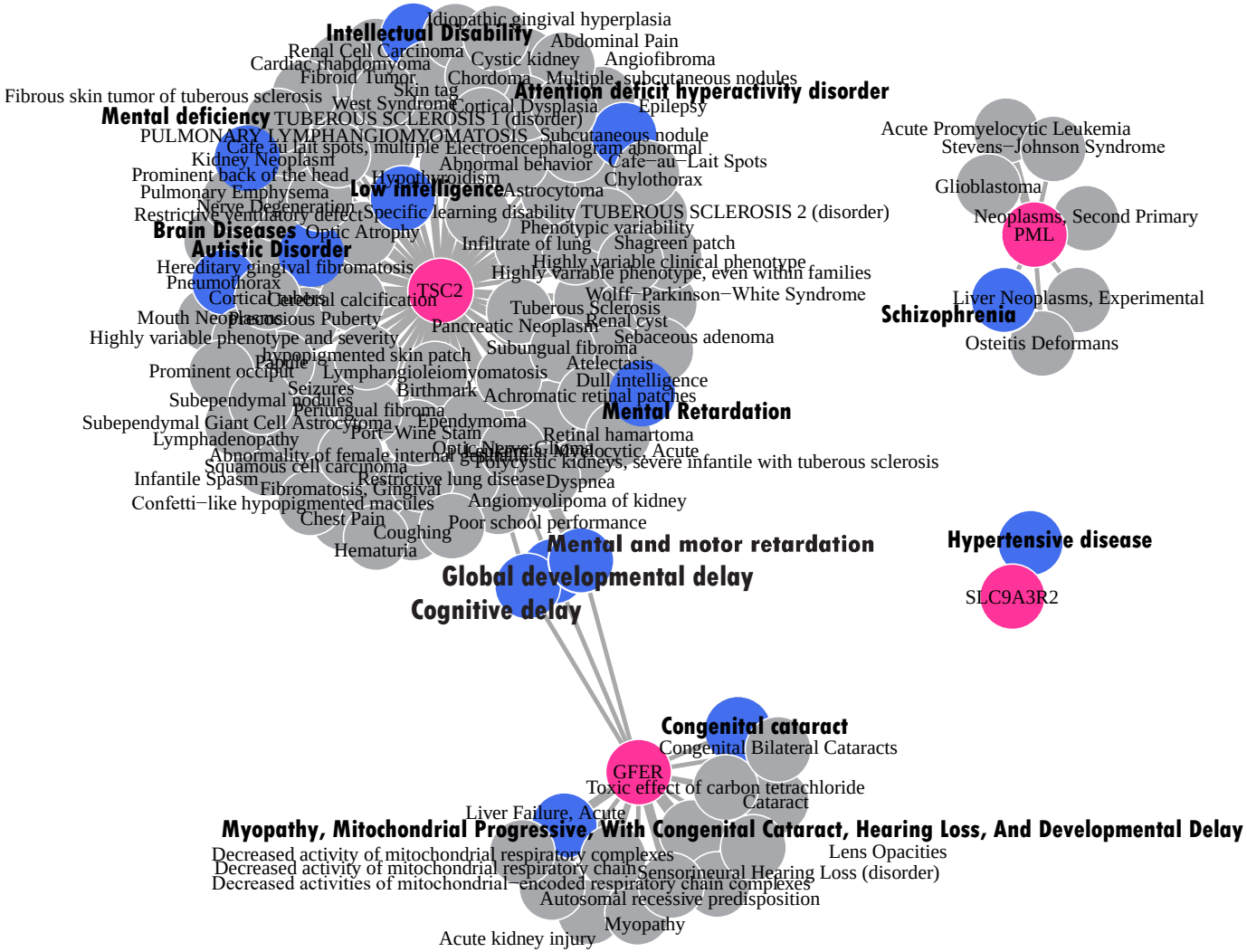


Figure 19: *Disease-gene network for genes identified for the association between control vs Alzheimer's disease with hypertension. The diseases (blue circles) overlapping between genes (pink circles) are bolded.*

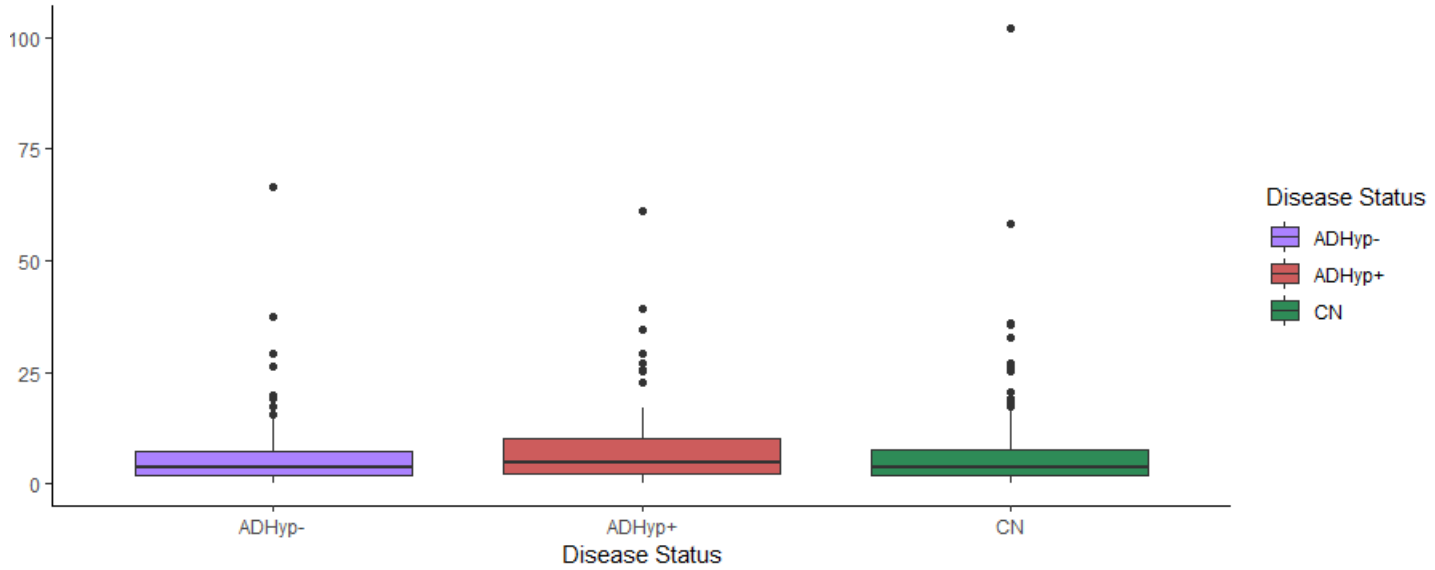
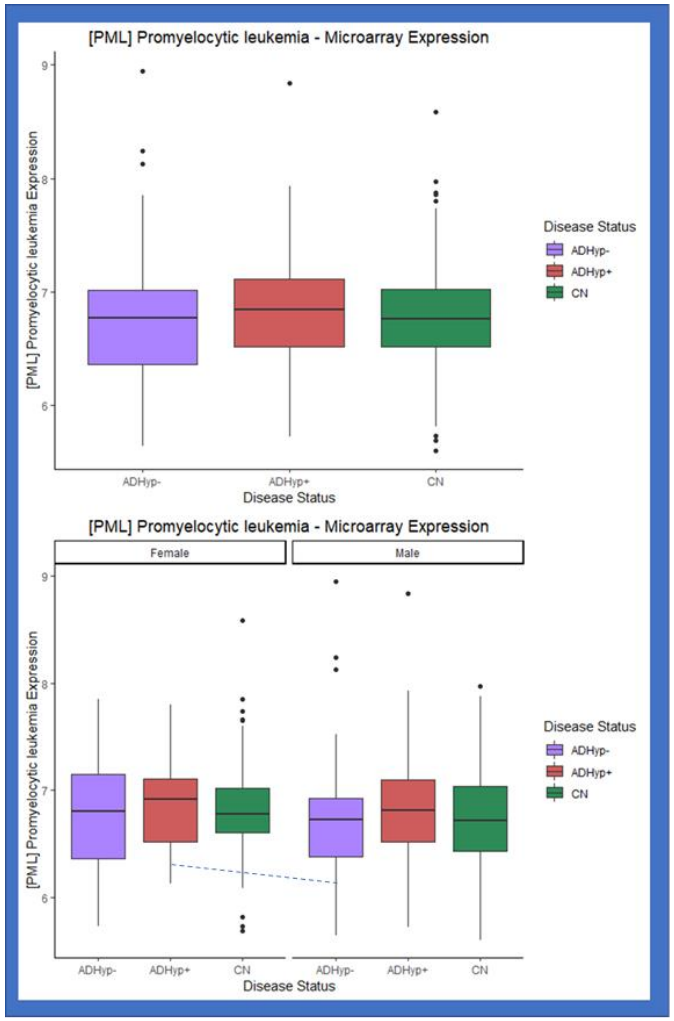


Figure 20: *Disease-based comparison of White Matter Hyperintensity (WMH) in ADNI-2/Go cohort (N=494) based on disease status – controls (CN), Alzheimer's disease without hypertension (ADHyp-) and Alzheimer's disease with hypertension comorbidity (ADHyp+).*



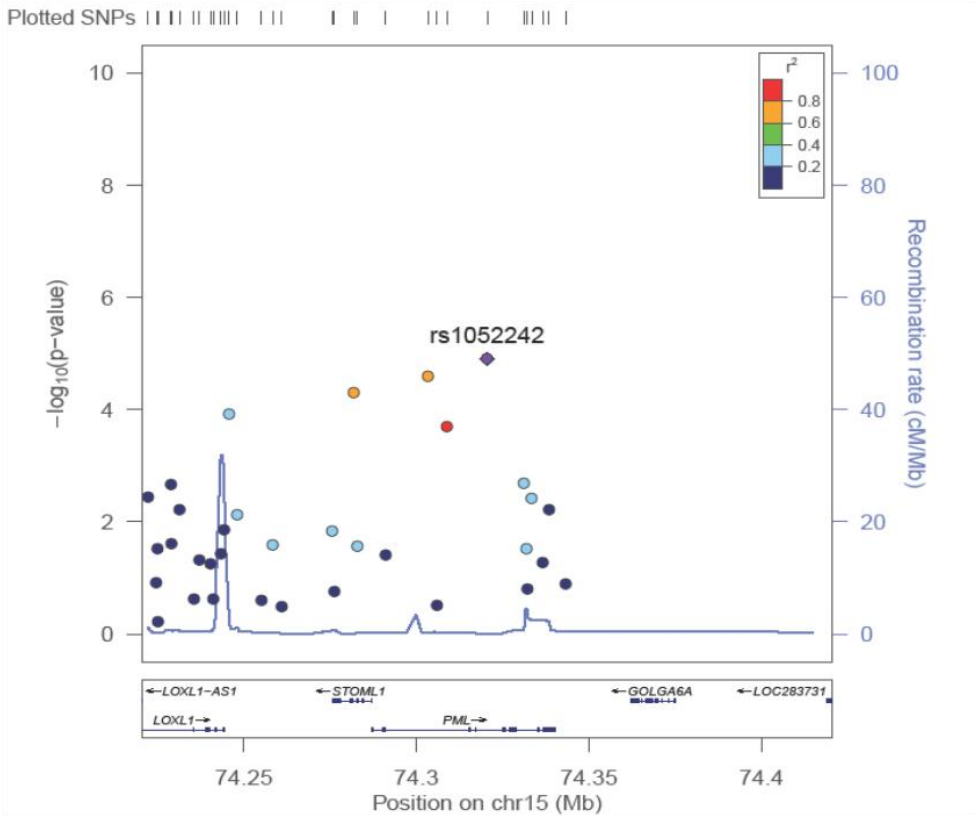
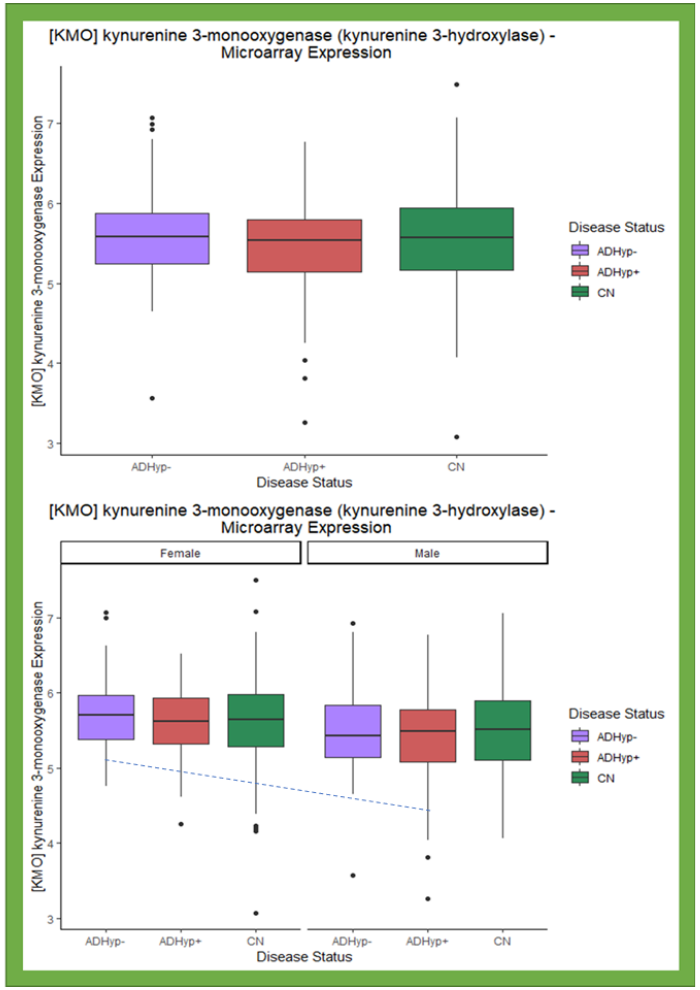


Figure 21: *The gene expression values of genes identified from the gene-based GWAS. (Left panel) Overall and sex-based differences in gene expression of PML gene in ADNI-2/GO cohort. Significant differences were seen for ADHyp-:Male-ADHyp+:Female, p = 0.046. (Right panel) The regional plot for rs1052242, which is the leading significant SNP for PML gene as observed in the gene based GWAS.*



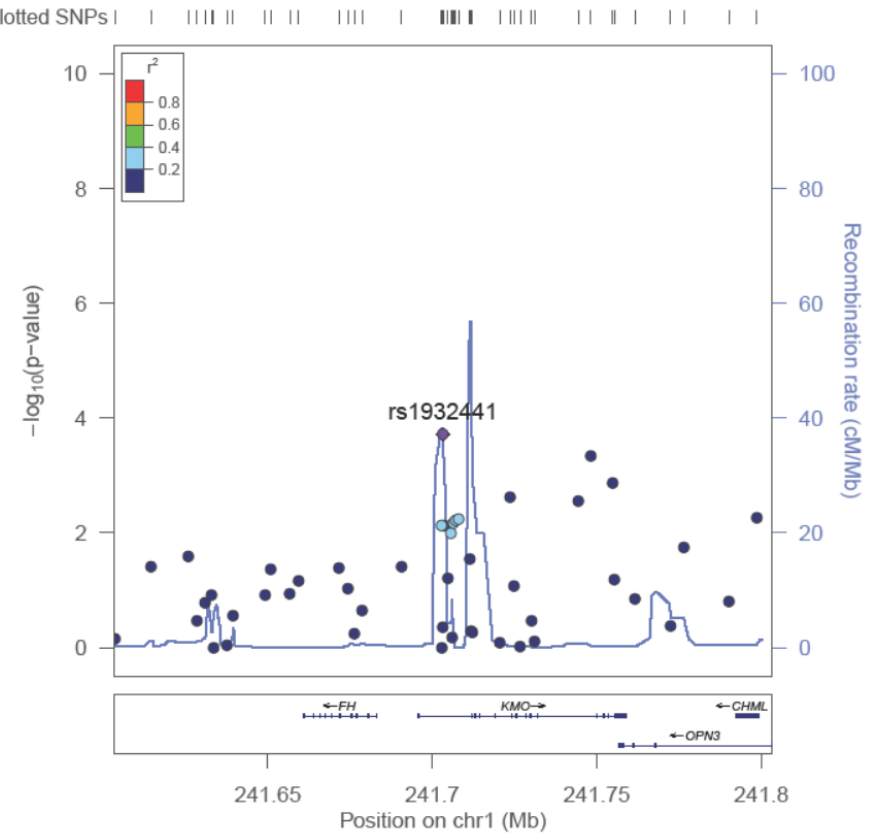


Figure 22: *The gene expression values of genes identified from the gene-based GWAS. (Left panel) Overall and sex-based differences in gene expression of KMO gene Significant differences were seen for ADHyp+:Male-ADHyp-:Female, p = 0.012. (Right panel) The regional plot for rs1932441, which is the leading significant SNP for KMO gene as observed in the gene based GWAS.*

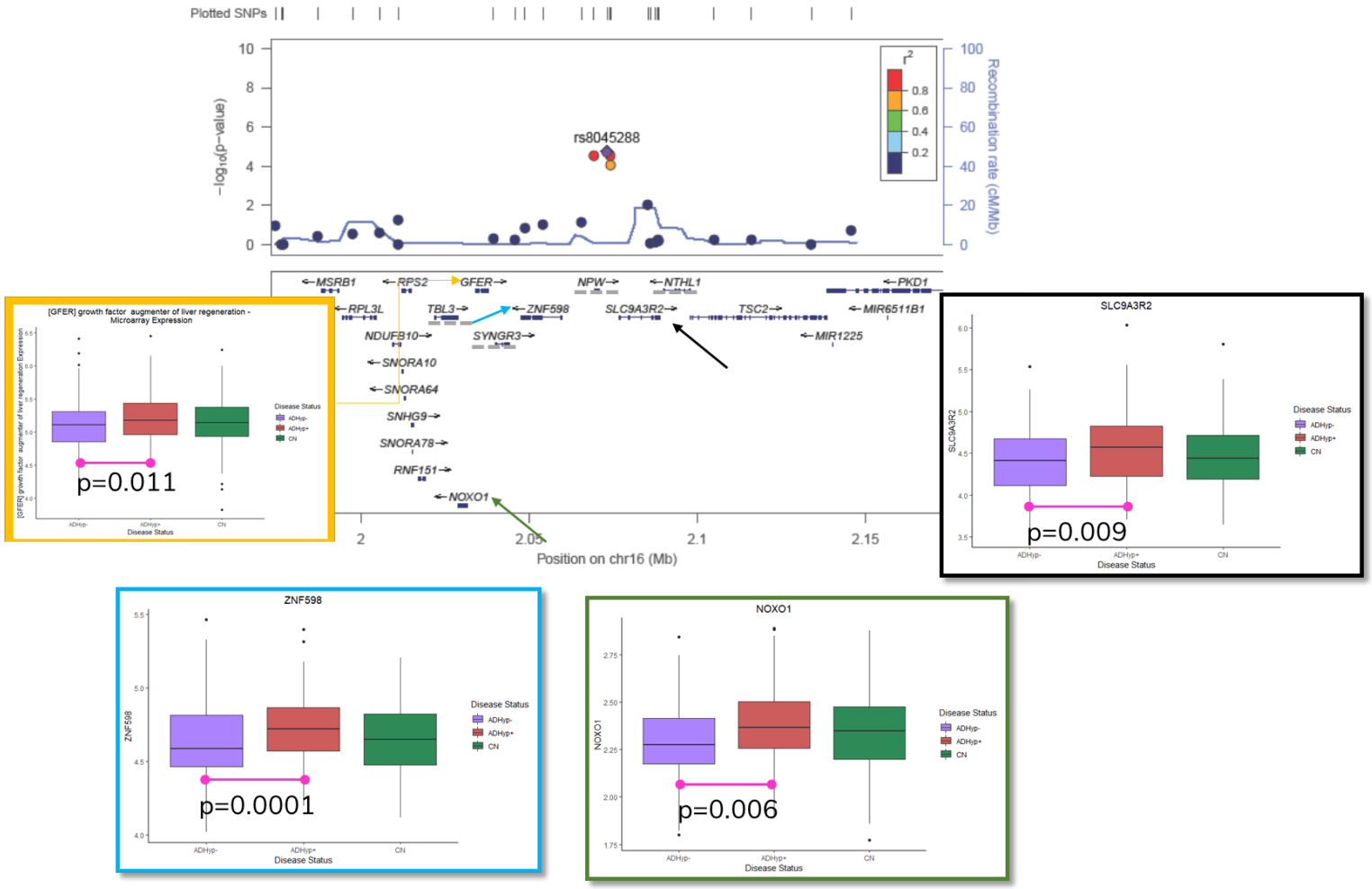


Figure 23: *The gene expression values of genes identified from the gene-based GWAS. The top panel shows the regional plot for rs8045288, which was annotated to multiple genes due to proximity (±50kb) from the SNP site. The gene expression of these genes was tested in ADNI-2/GO dataset as candidate genes affected by the SNP. Four genes – GFER, SLC9A3R2, ZNF598 and NOXO1 were found to have significant differences based on disease status. The gene expression values of other genes – NPW, NTHL1, TBL3, and SYNGR3, marked with gray dashed lines were not statistically significant between the three groups.*

3.5 REFERENCES

1 Mark Mather, L. A. J., Kelvin M. Pollard. Aging in the United States. (Population Reference Bureau, 2015).

2 Alzheimer's, A. 2011 Alzheimer's disease facts and figures. *Alzheimers Dement* **7**, 208-244, doi:10.1016/j.jalz.2011.02.004 (2011).

3 Emdin, C. A. *et al.* Blood Pressure and Risk of Vascular Dementia: Evidence From a Primary Care Registry and a Cohort Study of Transient Ischemic Attack and Stroke. *Stroke* **47**, 1429-1435, doi:10.1161/STROKEAHA.116.012658 (2016).

4 Kalaria, R. N. Neuropathological diagnosis of vascular cognitive impairment and vascular dementia with implications for Alzheimer's disease. *Acta neuropathologica* **131**, 659-685, doi:10.1007/s00401-016-1571-z (2016).

5 Smith, Eric E. Clinical presentations and epidemiology of vascular dementia. **131**, 1059-1068, doi:10.1042/CS20160607 %J Clinical Science (2017).

6 Modir, R., Gardener, H. & Wright, C. *Blood Pressure and White Matter Hyperintensity Volume—A Review of the Relationship and Implications for Stroke Prediction and Prevention*. Vol. 7 (2012).

7 Gregson, J. *et al.* Blood pressure and risk of dementia and its subtypes: a historical cohort study with long-term follow-up in 2.6 million people. *European journal of neurology*, doi:10.1111/ene.14030 (2019).

8 Bergantin, L. B. Hypertension, Diabetes and Neurodegenerative Diseases: Is there a Clinical Link through the Ca2+/cAMP Signalling Interaction? *Current hypertension reviews* **15**, 32-39, doi:10.2174/1573402114666180817113242 (2019).

9 Dugger, B. N. *et al.* A Cross-Sectional Analysis of Late-Life Cardiovascular Factors and Their Relation to Clinically Defined Neurodegenerative Diseases. *Alzheimer Dis Assoc Disord* **30**, 223-229, doi:10.1097/WAD.0000000000000138 (2016).

10 Paraskevas, G. P. *et al.* CSF biomarker profile and diagnostic value in vascular dementia. *European journal of neurology* **16**, 205-211, doi:10.1111/j.1468-1331.2008.02387.x (2009).

11 Saykin, A. J. *et al.* Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimers Dement* **11**, 792-814, doi:10.1016/j.jalz.2015.05.009 (2015).

12 Jostins, L. & McVean, G. Trinculo: Bayesian and frequentist multinomial logistic regression for genome-wide association studies of multi-category phenotypes. *Bioinformatics* **32**, 1898-1900, doi:10.1093/bioinformatics/btw075 (2016).

13 Liu, J. Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* **87**, 139-145, doi:10.1016/j.ajhg.2010.06.009 (2010).

14 Gutiérrez-Sacristán, A. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research* **45**, D833-D839, doi:10.1093/nar/gkw943 (2016).

15 Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research* **38**, W214-W220, doi:10.1093/nar/gkq537 (2010).

16 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).

17 Korb, E. & Finkbeiner, S. PML in the Brain: From Development to Degeneration. *Frontiers in oncology* **3**, 242-242, doi:10.3389/fonc.2013.00242 (2013).

18 Campesan, S. *et al.* The kynurenine pathway modulates neurodegeneration in a Drosophila model of Huntington's disease. *Curr Biol* **21**, 961-966, doi:10.1016/j.cub.2011.04.028 (2011).

19 Bartosiewicz, J. *et al.* The activation of the kynurenine pathway in a rat model with renovascular hypertension. *Experimental biology and medicine (Maywood, N.J.)* **242**, 750-761, doi:10.1177/1535370217693114 (2017).

20 Giil, L. M. *et al.* Kynurenine Pathway Metabolites in Alzheimer's Disease. *J Alzheimers Dis* **60**, 495-504, doi:10.3233/jad-170485 (2017).

21 Ibrahim, S. & Weiss, T. S. Augmenter of liver regeneration: Essential for growth and beyond. *Cytokine & Growth Factor Reviews* **45**, 65-80, doi:https://doi.org/10.1016/j.cytogfr.2018.12.003 (2019).

22 Kobayashi, K., Monkawa, T., Hayashi, M. & Saruta, T. Expression of the Na+/H+ exchanger regulatory protein family in genetically hypertensive rats. *Journal of hypertension* **22**, 1723-1730 (2004).

23 Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet* **51**, 51-62, doi:10.1038/s41588-018-0303-9 (2019).

24 Ma, M. W. *et al.* NADPH oxidase in brain injury and neurodegenerative disorders. *Molecular Neurodegeneration* **12**, 7, doi:10.1186/s13024-017-0150-7 (2017).

Chapter 4

TWO-STAGE BAYESIAN GWAS OF 9,638 INDIVIDUALS IDENTIFIES SNP-REGIONS THAT ARE TARGETED BY MIRNAS INVERSELY EXPRESSED IN ALZHEIMER'S AND CANCER

GITA A PATHAK, ZHENGYANG ZHOU, TALISA SILZER, ROBERT C BARBER, NICOLE R PHILLIPS

In revisions in Journal of Alzheimer' & Dementia (July 2019)

ABSTRACT.

INTRODUCTION: We compared genetic variants between Alzheimer's disease and two age-related cancers – breast and prostate cancer– to identify variants (SNPs) that are associated with inverse comorbidity of AD and cancer.

METHODS: Bayesian multinomial regression was used to compare sex-stratified cases (AD and cancer) against controls in a two-stage study. A ±500 KB region around each replicated hit was imputed and analyzed after merging individuals from the two stages. The miRNAs that target the genes involving these SNPs were analyzed for miRNA family enrichment.

RESULTS: We identified 137 variants with inverse odds ratios for AD and cancer located on chromosomes 19, 4 and 5. The mapped miRNAs within the network were enriched for miRNA-17 and miR-515 families.

DISCUSSION: The identified SNPs were rs4298154 (intergenic), within TOMM40/APOE/APOC1, MARK4, CLPTM1, and near the VDAC1/FSTL4 locus. The miRNAs identified in our network have been previously reported to have inverse expression in AD and cancer.

Short Title: Bayesian GWAS of inverse comorbidity between Alzheimer's and cancer

4.1 INTRODUCTION

The aging population, defined as 65 years or older, is expected to experience a substantial demographic shift. By the year 2060, this group is expected to reach 98 million in US, placing an unprecedented burden on the healthcare system [1]. Due to the role of aging in accumulation of physiological deterioration, the number of chronic diseases affecting this population continues to rise and many individuals suffer the co-occurrence of two or more diseases (i.e. comorbidity) [2]. In contrast to the more common direct coexistence of diseases, some chronic age-associated diseases have been identified to be inversely comorbid – a lower-than-expected occurrence of the secondary disease after the index/first diagnosis [3]. These intriguing inverse associations between select diagnoses have garnered much attention in the last few years, as they shed light on the heterogeneity of age - associated multifactorial diseases.

Alzheimer's disease (AD) and cancer are two predominant age-associated diseases which are inversely comorbid as reported by several epidemiological findings. In a meta-analysis of association studies from 1966 to 2013, AD patients had a decreased incidence of cancer by 42% and individuals with cancer history had 37% reduced risk of Alzheimer's disease [4]. In a recent retrospective study of more than 3 million US veterans, cancer survivors aged ≥65 years had a lower risk of AD than other age-related outcomes. The odds ratio was 0.89 in fourteen cancer types after excluding melanoma, prostate, and colorectal cancer [5]. These findings have fueled numerous exploratory investigations into possible genetic mechanisms that may be responsible for this inverse association between two common age-related diseases.

While genome wide association studies (GWAS) have identified multiple genetic loci contributing to either AD or cancer, no study has reported cross-phenotypic effects of individual genetic

variants [6-8]. Therefore, we sought to detect variants that confer inverse risk for AD and cancer by (1) harmonizing the intermediate risk factor – age, between the two disease populations, and (2) directly comparing cases which represent the two extremes of the phenotypic variance to a common set of controls. The comparison of multiple cases to controls (also known as cross-disorder studies) warrants the use of multinomial regression. Multiple GWAS scenarios such as fine-mapping of variants[9], pleiotropic and regulatory variants employ Bayesian methods in genome-wide studies to achieve higher accuracy and prediction than frequentist approach[10]. The use of priors in Bayesian approach allows for investigation of multiple directionalities among phenotypes [11]. Here, we used flat/ default priors [12] to test the relationship between AD and cancer in a conservative setting of no SNP effect exists between the two diseases[11]. We aimed to address this goal by conducting a Bayesian multinomial GWAS (B-GWAS) to identify genetic variants that confer inverse risk in the aging population between 60-80 years for AD and cancer, using the two most prevalent age-related cancers: breast cancer and prostate cancer [13]. We conducted a B-GWAS in two phases – discovery and replication, comparing AD and cancer to common controls. All data sets were stratified by gender and harmonized on age. Replicated hits were further investigated in the merged data set (discovery and replication) via B-GWAS of imputed genotypes within 1Mbp window (± 500 KB) of each hit, and conditional analysis was used to identify secondary hits.

4.2 METHODS AND MATERIALS

4.2.1 Data description

We obtained access to datasets from Alzheimer's Disease Genetics Consortium (ADGC) (phs000372.v1.p1) [14] and Breast Prostate Cancer Care Consortium (BPC3) (phs000812.v1.p1) [15] via dbGaP's authorized application to individual level genotype data. We also obtained access

to Alzheimer's Disease Neuroimaging Initiative (ADNI) (www.adni-info.org) (see Supplementary file3 – Text S1 for details). These datasets were chosen because they all were genotyped on the same genotyping platform – Illumina Human660W-QuaSd, to minimize any technical bias and harmonization issues while merging the datasets and potential array specific inaccuracies during imputation. The research protocol for this project was reviewed by University of North Texas Health Science Center Institutional Review Board on June 24, 2016 and determined to be exempt human subject research under IRB–2016-090.

The first cohort in the BPC3 dataset [15] - phs00812.BreastProstateCancer.v1.p1.c1 (BPC3-c1) - had a total of 4915 participants. There were 2314 individuals in the prostate cancer group with genotype data for 583132 SNPs. The total number of individuals in the breast cancer group was 2601 with genotyped data of 541219 SNPs. The second cohort in BPC3 - phs00812.BreastProstateCancer.v1.p1.c4 (BPC3-c4) - had 4664 participants. The total number of individuals in the prostate cancer group in this cohort was 4069 with genotype data for 583132 SNPs. The total number of individuals in the breast cancer group was 595 with genotyped data of 541219 SNPs on hg18 build. The ADGC dataset [14] had genetic variant data of 6065 individuals genotyped on the Illumina platform via Human660WQuad array. The ADC1 dataset had 2905 individuals with 657366 genotyped markers, and the ADC2 dataset had 1170 individuals with 657366 genotyped markers on hg19 build. We combined these two datasets for the final ADGC dataset. From the ADNI cohort, we used the ADNI-1 dataset as it was also genotyped on Illumina Human660W-Quad array. The ADNI-1 dataset had 757 individuals with 620902 typed SNP markers.

4.2.2 Merging and Quality Control of datasets

For the discovery stage, we merged individual level genotype data from ADGC and BPC3 (c1) by aligning strand and genomic build to hg19 using PLINK (v1.9) [16]. A total of 8990 individuals (self -described European ancestry) remained in the merged dataset with 539774 common genotyped SNPs. We then followed quality control protocol outlined by Anderson et al.[17], principal components were calculated using R package – 'SNPRelate' [18]. Details of QC protocol and number of individuals removed at each stage are outlined in Supplementary file3 FigureS2.

For the replication stage, we merged individual level genotype data from ADNI and BPC3 (c4) by harmonizing strand using the array manufacturer's documentation and converted to genomic build hg19. After QC, we were left with 486308 markers in 4226 individuals. This QC'd dataset was stratified on sex, with 3580 male and 646 females.

4.2.3 B-GWAS

For the discovery stage, we conducted genome-wide Bayesian multinomial logistic regression (B-GWAS) adjusted for age, and ancestry - PCs 1-10 (self-described as Caucasian/European ancestry) using Trinculo [11] with the default prior parameter. The variance proportion for each first three PCs was less than 2% for each dataset. There was no significant difference ($p > 0.05$) in means of eigenvalues between phenotypes after first three PCs; calculated in EIGENSTRAT [19] for both stages of datasets. We compared 900 males with AD and 997 males with prostate cancer to a combined male control population of 802 individuals from the two datasets. Similarly, we conducted B-GWAS in females, comparing 946 females with AD and 460 females with breast cancer to a combined female control population of 1307 individuals. We then selected SNPs that were significant with a joint log Bayes factor ≥3 and odds ratio in the opposite direction for the two diseases (OR_Alzheimer's disease > 1 & OR_Breast/Prostate <1 or OR_Alzheimer's disease

< 1 & OR_Breast/Prostate > 1). Next, in a replication phase, these top significant variants were analyzed in independent sample sets. Consideration of significance threshold and comparison to P-values was motivated by works of Dr. Wakefield [20, 21].

For the replication stage, we followed the same QC protocol, comparing 149 males with AD, and 1046 males with prostate cancer, to a combined control male population of 2385 individuals. Correspondingly, association analysis was performed in 107 females with AD and 229 females with breast cancer to a combined female control population of 310 individuals.

4.2.4 Regional GWAS

Imputation was performed using IMPUTE2 [22] with 1000 genomes Phase 3 dataset for each of significant hit regions. The datasets from the two stages were merged for a 1Mb region (± 500 KB) of the replicated hits that were identified in the B-GWAS. Following SNP-level QC, a regional GWAS was performed in 6279 males comparing AD and prostate cancer against controls on chromosome 4. Similarly, following QC, regional GWAS was performed in 3359 females comparing AD and breast cancer against common controls on chromosome 19 and 5. For the regional GWAS on chromosome 19, we identified haploblocks of the 1Mb region using plink v1.9. There were four TOMM40 haploblocks of 21, 6, 2 and 11 SNPs and one APOE haploblock containing 6 SNPs. To identify higher risk SNPs between TOMM40 and APOE, a total of four separate haploblock-based associations were conducted merging each of the TOMM40 and APOE haploblocks, keeping only individuals with all SNPs. To identify secondary hits on chromosome 19, conditional tests were performed using APOE SNPs – rs429358 and rs769449. For other regions, the top two significant SNPs were used to perform conditional analysis.

4.2.5 SNP annotations and testing for enriched processes

All significant SNPs were mapped to genes using Ensembl's Ch37 Variant Effect Predictor (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP), and functional annotations were retrieved from SNP nexus [23]. Figures were plotted using ggplot2, gene annotations were visualized with Gviz package, and LD map was created using LDheatmap package in R. The mapped genes were tested for functional and diseases processes and visualized using Ingenuity Pathway Analysis® (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis).

4.2.6 miRNA enrichment analysis and text mining for miRNA expression

All the annotated significant genes were used as input in miRNet [24]. The miRNA nodes in network were filtered on degree filter of 1.0, to reduce orphan miRNAs. The filtering prioritizes miRNAs with at least two connecting query genes. The miRNAs in the network were then assessed for miRNA family enrichment using hypergeometric test, with a p-value less than 0.05 considered as significant.

4.3 RESULTS

4.3.1 GWAS in sex-stratified Alzheimer's disease and cancer

We found 422 SNPs to be significant as shown in the bokeh plot (Figure 1, top panel) by comparing control males with males with AD and males with prostate cancer (Supplementary file1 TableS1). Similarly, we tested the relationship of AD and breast cancer, by comparing females with AD, and females with breast cancer against female controls, which resulted in 324 significant loci (Supplementary file1 TableS3) with inverse odds ratios between AD and breast cancer (Figure 1, bottom panel).

We evaluated these top significant SNP loci in another dataset of individuals, merging Alzheimer's population of ADNI-1, and breast & prostate cancer population of BPC3 (c4) followed by QC procedures. The QC'd dataset was then separated by sex. In this replication stage, out of 422 significant SNPs from the discovery stage, 411 SNPs were present in the replication dataset (Figure 2, top panel). The association analysis, adjusted for age and PCs 1-10, revealed one SNP that replicated in this dataset - SNP_allele : rs4298154_C an intergenic SNP on chromosome 4, had the odds ratio of 0.77 for AD (logBF – 1.9) and 1.251 for prostate cancer (logBF – 3.7) with an overall logBF of 5.14 (Supplementary file1 TableS2).

In females, out of 324 significant SNPs from the discovery stage, 309 were present in the replication stage. The association test replicated two significant loci (Supplementary file1 TableS4) : a) rs2075650_A had odds ratio of 0.52 for AD (logBF – 13.418) and 1.245 for breast cancer (logBF – 3.17), with an overall logBF of 14.458 ; b) rs17700949_A, mapped on chromosome 5 near *C5orf15* and *VDAC1* gene, had odds ratio of 0.817 for AD (logBF – 1.74) and 1.256 for breast cancer (logBF – 2.483), with an overall logBF of 3.33 (Figure 2, bottom panel).

4.3.2 Regional GWAS of replicated SNP hits and Conditional analysis

In order to achieve finer SNP resolution, a 1Mb region was imputed around the replicated hits (± 500 KB), here after referred to as 'risk region' in each dataset. The two datasets for this risk region analysis were merged to improve power by increasing sample size. Following SNP-level QC, a regional B-GWAS with default prior and adjustment for age and PCs 1-10 from the merged datasets was performed in 6279 males comparing AD and prostate cancer against controls on chromosome 4. For the risk region in chromosome 4, a total of 3107 variants were analyzed by B-GWAS and 21 SNPs were found to be significant (Supplementary file1 TableS5), the most

significant SNPs were a) rs4298154_C having odds ratio of 0.813 and 1.223 for AD and prostate cancer, respectively, with overall logBF of 10.68, and b) rs57139228_G having odds ratio of 0.885 and 1.156 for AD and prostate cancer, respectively, with overall logBF of 5.63 (Figure 3). After conditioning on these hits, no SNPs remained significant.

Similarly, regional B-GWAS was performed in 3359 females comparing AD and breast cancer against common controls on chromosome 19 and 5. Of 1499 variants in the risk region on chromosome 19, 113 SNPs were found to be significant (Supplementary file1 TableS6). The top significant SNPs were a) rs34404554_C having odds ratios of 0.321 and 1.199 for AD and breast cancer, respectively, with overall logBF of 181.5, b) rs71352238_T with odds ratio of 0.321 and 1.206 for AD and breast cancer, respectively, with overall logBF of 180.83, and c) rs2075650 having odds ratio of 0.325 and 1.18 for AD and breast cancer respectively with overall logBF of 177.63. All three SNPs were mapped to *TOMM40*. Select top significant SNPs are labelled in Figure 4 (left panel). To identify which SNPs between *TOMM40* and *APOE* were most significant, we conducted association with SNPs in each haploblock of *TOMM40* and *APOE* keeping individuals who had complete set of SNPs. Here, we found the same results: *APOE* SNPs were more significant than *TOMM40*, and after conditioning on the top two *APOE* SNPs, none of the *TOMM40* were significant for both diseases; significance was observed only for Alzheimer's disease (logBF > 1.5). Since the *TOMM40* SNPs were not independent of *APOE* we conditioned the association analysis using *APOE* SNPs – rs429358 and rs769449. After conditioning, none of the SNPs were significant for both diseases.

The second replicated hit between AD and breast cancer was in chromosome 5, we analyzed 2340 SNPs in this 1Mb region for inverse association. A total of three SNPs remained significant including the replicated hit a) rs17700949_A having odds ratios of 0.875 (logBF – 4.09) and 1.175

(logBF – 4.54) for AD and breast cancer, respectively, with overall logBF of 6.09, b) rs10068691_G having odds ratios of 1.13 (logBF – 2.64) and 0.88 (logBF – 2.2) for AD and breast cancer, respectively, with overall logBF of 3.19, and c) rs1109309_G having odds ratio of 0.905 (logBF – 2.05) and 1.14 (logBF – 2.77) for AD and breast cancer, respectively, with overall logBF of 3.15 (Figure 4 ; right panel) (Supplementary file1 TableS7). All three SNPs are in close proximity to one another and mapped to an intergenic region between *FSTL4* and *VDAC1* based on GRCh37/hg19.

The significant SNPs on chromosome 4 are in the intergenic region, and the nearby genes are *ARAP2, DTHD1*, and *KIAA1239*. Additionally, annotation was retrieved from the Genetic Association Database using SNP-nexus, which showed association of these SNPs with *HTRA3*, *AREG*, and *NRAS* (Supplementary file2). Interestingly, some of the variants in this region had a slightly higher CADD score: rs4565101 had a score of 7.741, and rs58262946 at 6.342; suggesting that these SNPs are deleterious.

In the regional B-GWAS of AD and breast cancer on chromosome 19, the significant SNPs mapped to *PVRL2, CTB-129P6.4, TOMM40, APOE, APOC1, APOC2, APOC4, APOC4-APOC2, CTB-129P6.11, CLPTM1, RELB, AC005779.1, MARK4, AC006126.3*, and *AC005779.2*, and their functional annotation identified *BCAM, ZNF107, ZNF92* and *ZNF138* (Supplementary file2). SNPs in the chromosome 5 risk region are intergenic to *FSTL4* and *VDAC1*. Other genes within 300 kb include *TCF7, SKP1, PPP2CA, CDKL3* and *UBE2B*. The top significant SNP - rs17700949 had CADD score of 4.224, GWAVA score of 0.53 and ReMM score of 0.603, indicating a slightly deleterious effect.

4.3.3 Gene network analysis using Ingenuity Pathway Analysis

We analyzed the query genes to test for enriched processes using IPA's biobase knowledge. Some of the represented processes were inflammatory and cellular interactions, including LXR/RXR activation, Wnt/β-catenin, PI3K/Akt and sirtuin signaling pathway (Supplementary file3 FigureS3). The query genes resulted in two networks (Supplementary file3 FigureS4), we merged the networks and examined for leading canonical pathways (Figure 5).

4.3.4 miRNA annotation and enrichment analysis

We used the query genes to identify interacting miRNAs, the network was constructed using miRNet [24] (Figure 6). All the miRNAs in the network were assessed for miRNA family enrichment; we observed miR-515 (15 members) and miR-17 family (6 members - miR-17, 20a, 20b 106a, 106b and 93) to be significant (Supplementary file3 TableS2). The miRNAs with the highest number of interactions (Supplementary file3 Table S3) were investigated to find the direction of their expression changes in AD and cancer in the same tissue wherever possible. Intriguingly, when comparing their expression direction in the same tissues, we see an inverse direction for these miRNA expression levels in the two diseases. Examining for miRNA-SNP binding site, we observed two SNPs – rs6859 and rs11556505 to disrupt binding sites for multiple miRNAs (Supplementary file3 TableS4 & S5).

4.4 DISCUSSION

Several GWAS have been conducted for Alzheimer's disease, breast and prostate cancer and recently a few meta-analyses have investigated the relationship between AD and cancer. Feng et al. (2017), Sanchez-Valle (2017) and Ibanez (2014) have reported that the genetic relationship between AD and cancer varies based on cancer subtypes. This is the first study that focuses on

analyzing cross-phenotypic differences of SNPs using individual-level genotype and targeting two most common age-associated cancers – breast and prostate cancer.

4.4.1 Genic context of SNP hits in Alzheimer's disease and cancer

Since the top hits for breast cancer were in the *TOMM40/APOE* region, we conducted separate association analysis for SNPs in the *TOMM40* and *APOE* haplotypes among individuals with no missing SNPs to determine if *TOMM40* was independent of *APOE*. After conditioning on the two *APOE* SNPs – rs429358 and rs769449, we found that *TOMM40* SNPs were either marginally significant for AD, or insignificant for both AD and cancer. Therefore, our findings indicate that the cross-phenotypic effects of SNPs exist in the *TOMM40/APOE* region. Multiple GWAS have reported *APOE*'s association with risk for AD [25]. Intriguingly, *APOE* also has been a subject of investigation for carcinogenesis. A meta-analysis conducted by Anand et. al [26] reported a negative association between *APOE4+* genotypes and the overall risk for cancer subtypes. Studies have reported that genotyping *TOMM40*'523 loci leads to a better prediction of AD over *APOE* predictions alone [27]. In cancer, *TOMM40* expression surface antigens were elevated in pancreatic cell lines, and gene expression was upregulated in ovarian cancer cell lines [28, 29]. This heterogeneous phenotypic association of *TOMM40/APOE/APOC1* region is evinced in our association for AD and cancer and is also summarized by Yashin et. al [30].

In the chromosome 19 region, we also identified SNPs within *MARK4* exhibiting an inverse relation between the two diseases. *MARK4* belongs to the microtubule affinity-regulating kinases family. *MARK4* and its family play role in phosphorylation of tau, mediated by co-expression of *APP* (Amyloid Precursor Protein) and *MARK/Par-1*, as evidenced by their phosphorylated products in granulovacuolar bodies in brain tissues of AD patients [31]. In cancer, elevated *MARK4* expression is found to be correlated with cancer severity in breast, lung and prostate

neoplasms [32] mediating via the upregulation of the Wnt signaling and negative regulation of mTORC. Elevated *MARK4* has also been shown to stimulate tumorigenic properties in breast cancer cells by impeding Hippo signaling [33]. Another neighboring gene – *CLPTM1* (Cleft Lip and Palate Associated Transmembrane Protein 1) in this region was recently found to be independently associated with AD in GWAS-derived expression study [34]. An interesting finding here was that *CLPTM1L/CRR9*, a paralogue of *CLPTM1* is widely attributed as a risk-factor for multiple cancer subtypes as informed by genome-wide association and experimental studies [35-37]. Inoue et. al detected elevated expression of both *CLPTM1* and *CRR9* in oral squamous cancer cells [35]. *APOC1* in this region is associated with the formation Aβ plaques in Alzheimer's disease and is under expressed in subjects with *APOE4* genotype [38]. Conversely, *APOC1* is overexpressed in cancer tissues and influences the MAPK pathway triggering cellular expansion and motility [39]. *RELB* expression is correlated with tumor development and inflammatory processes [40] and the cumulative effect of rare variants in *RELB* is associated with amyloid burden in the cortical region of AD patients [41].

In the chromosome-4 risk region, the top variants are intergenic and the nearest pseudogene *SEC63P2* (SEC63 Homolog (*S. Cerevisiae*) Pseudogene 2) – has been associated with coronary artery calcification [42] and body mass index [43]. The closest (~850KB) coding gene to this region is *ARAP2*. Expression of *ARAP2* also has been reported as part of a risk score prediction for pancreatic cancer by Liu et. al [44]. Variants in *ARAP2* are known to be *TP53* binding sites, and are associated with advanced prostate cancer [45]. Experimental studies have highlighted the role of *ARAP2* in cytoskeleton remodeling and axonal transport mediated by neurotoxin-stress and dysregulating motor neurons [46].

Other loci implicated in inverse risk for breast cancer and AD are mapped between *FSTL4* and *VDAC1* on chromosome 5 using GrCh37 build. *VDAC1* – Voltage Dependent Anion Channel 1 - is a key mitochondrial-mediated apoptotic protein that acts via BCL-2 pathway, *VDAC1* also interacts with *TOMM40* for mitochondrial transport in the *PINK-1/PARK* pathway. In AD brain tissue, higher *VDAC1* expression is found in neurites with Aβ deposits [47]. In cancer, metabolic reprogramming has been attributed to *VDAC1*, and its apoptotic properties have become a pharmacological target of interest [48]. The upstream gene to our significant variants was *FSTL4* – Follistatin Like 4 - known for its role in extracellular calcium ion binding. Interestingly, in the latest genome build – GrCh38, our significant variants are mapped within the *FSTL4* gene. Genome-wide studies have reported variants in this gene to be associated with lung carcinoma [49], cognitive impairment [50, 51] and hypertension [52].

Overall, the genes containing the cross-phenotypic SNPs have known pathological roles in both AD and cancer.

4.4.2 Role of enriched processes in Alzheimer's disease and cancer

The enriched biological processes involved with our query genes were sirtuin pathway, Wnt-signaling, LXR-RXR and PI3K/Akt mechanism. Our findings are consistent with Ibanez et. al, who reported dysregulation in Wnt-signaling - upregulation in cancer, and downregulation in neurological diseases. Wnt-signaling is implicated in metastatic cell proliferation [53]. The PI3K/Akt signaling is associated with metabolic dysfunction inducing insulin stress via deregulation of insulin receptors. While cancer cells are suspected to thrive on glycolytic byproducts from insulin stress, the brain is affected by the disturbances in PI3K/Akt signaling and exhibits cognitive deficits [54]. LXR-RXR are a class of transcription factors that also affect metabolic activity by regulating lipids and inflammatory responses [55]. Their expression in AD

animal models has been associated with cognitive deficits and increased Aβ levels in cerebrospinal fluid [55]. In cancer, LXR ligands interact with both Wnt and AKT signaling pathways and induce pyroptosis – inflammation-induced cell death [56]. Sirtuin proteins are involved in both cancer and Alzheimer probably due to their involvement in regulating mitochondrial biogenesis, interacting with *TOMM40* and *VDAC1*, and can detect peripheral metabolic dysregulation [57]. Altogether, the observed processes seem to regulate metabolic activities of metastatic cellular expansion and accumulation of amyloid burden in AD and cancer.

4.4.3 Inverse expression of miRNAs and their role in Alzheimer's disease and cancer

To investigate the potential mechanisms that underlie the observed genetic associations with inverse risk of AD and cancer, we turned our focus to genetic-based regulatory system – miRNAs [58].This provided the rationale for asking the following question: could the inverse effect of the genetic variants be due to miRNA mediated differential regulation of our candidate genes?

miRNAs are small non-coding RNA molecules that target multiple regions in a gene and are expected to regulate ~60% of transcripts, thus altering cellular metabolism. Due to their exosomal packaging, they allow cross-talk through the blood brain barrier and interact with other organs [59]. The reported genes are primarily targeted by select miRNAs including the miRNA-17 and 515 cluster, 125b, 335, and 26b which are differentially expressed in Alzheimer' disease and cancer (Table 1). This SNP-miRNA relationship highlights the role of these miRNAs in altering regulation of target regions identified from our study in Alzheimer's disease. The miRNA-17 cluster (including miRNA-106, 20 and 93) has been reported to regulate *APP* expression in brain and neuronal cells of sporadic AD patients [60]. Additionally, miRNA-106 is downregulated in the frontal cortex of AD patients, which increases Aβ1-42 and induces tau phosphorylation [61]. As underscored in our results (Table1), members of this 'onco-miR' [62] family show inverse

expression in the setting of AD versus cancer. MiR-515 family was also found to be enriched in our network. The upregulation of miR-515 in cancer is inversely correlated with survival of cancer patients [63]. The miR-515 suppresses p21 which is required for inducing senescence, which is also modulated by miR-106b's overexpression [64]. MiR-515 has been reported to be downregulated in the temporal cortex of AD patients [65]. The upregulation of miR-125b in multiple regions of the AD brain has been known to correspond with neurofibrillary tangles, primarily in gray matter region of post-mortem AD brains [66]. In human neuronal cells, activation of NF-κB pathways from deposition of Aβ results in overexpression of miR-125b [67]. On the other hand, 125b is under expressed in cancer cells which initiates cancer hallmarks [68]. 125b has also been observed to target BCL-2 and increase its apoptotic activity via BMF in Alzheimer's disease [66]. miR-335 is found to be under expressed in multiple cancer, which is regulated in a cyclical mode by p53 [69]. In contrariety, upregulation of miR-335 triggers p21 and lowers p53 expression which leads to increase in tau levels in AD patients [70, 71]. In mouse model of AD, miR-335 was overexpressed in hippocampus and lowering its expression was demonstrated to reduce cellular cholesterol and alleviate cognitive impairment [72]. The upregulation of miR-26b triggers expression of cyclin-dependent kinase 5 which initiates phosphorylation of tau and apoptosis in AD [73]. Conversely, increasing 26b results in anti-tumorigenic properties [74]. Multiple cancer types have been found to have under expression of miR-26b [75]. miR-34a has been reported to be over expressed in brain regions of AD possibly resulting in dysregulation of synaptic and metabolic activity [76]. The dysfunction of p53 governs the expression of miR-34a [77]which is deficient in most cancers[78]. In animal models of APP and presenilin 1 knockout, lowering miR-34a mitigates cognitive symptoms of AD [79].

Overall, the identified miRNAs play a dominant role in AD pathology and their targets reported here warrant functional studies to characterize their sequence-specific multi-gene regulation.

4.4.4 Role of SNPs in disrupting miRNAs and their target binding sites

Our analyses provide evidence of potential connections between miRNAs and genetic variants resulting in bidirectional effects between AD and cancer. Studies have shown that variants present in the 3'UTR (untranslated region) of mRNAs, known as poly-miRTs, change the half-life of mRNAs and thus alter protein expression [80]. Two of our reported cross-phenotypic SNPs – rs6859 (*PVLR2/NECTIN2*) and rs11556505 (*TOMM40*) alter miRNA binding sites. SNP rs6859 is found to alter binding site for miRNAs- 143, 147, 199, 584 and 648. The elevation of miRNA-143 reduces glucose uptake and promotes cellular apoptosis in cancer cells [81]. However, in AD, the elevation of miR-143 is localized in neurons and is proportional to Braak stages of neurofibrillary tangles in locus coeruleus region of AD brains [82]. Other miRNAs- 147, 199 and 584 are understudied in AD, but share a common function of suppressing the tumor and inhibiting cancer progression [83-85]. This demonstrates that a single nucleotide variation can alter sites for multiple miRNAs leading to variability in gene expression. Using polymiRTs database [86], rs11556505 is documented to modify miRsite for miR-484 (Supplementary file3 Table S5). Among several targets, miR-484 is reported to inversely alter *Fis1* expression that promotes mitochondrial fission[87]. *Fis1* interacts with sirtuin complex to trigger cell migration and invasion, and is over expressed in cancer [88]. In AD-brain derived fibroblasts and hippocampal tissue, mitochondrial fission proteins including *Fis1* is upregulated [89]. While the exact underpinnings of the interaction of *Fis1* on the reported *TOMM40* site are unknown we observe a common thread of mitochondrial dysregulation and sirtuin signaling leading to neuronal dysfunction and cancer expansion. Therefore, this site necessitates functional studies using whole-

genome and RNA sequencing to evaluate allele-specific expression of the reported SNP-miRNA-target site.

In conclusion, the induction of SNPs within and around these UTR sequences can have multiple functional consequences by altering miRNA binding sites, generating multiple transcripts which may be differentially targeted by miRNA regulators [90]. Variants in mRNA binding sites are relatively more common than variants in genes encoding miRNAs [91]. Our analysis identified SNPs that are indicative of causing possible perturbations in miRNA binding sites. MiRNAs, either acting individually or in combination, can thus result in differential transcriptional regulation of multiple genes (Supplementary file3 Figure S5). While we don't know the exact mechanisms that lead to perturbations in gene expression from intergenic SNPs, studies have shown that SNPs in non-coding regions such as introns, lncRNAs, intragenic regions can affect miRNA expression levels [92]. Recent studies have identified the importance of these reported miRNAs either via literature-driven or meta-analysis studies [93, 94]. However, by studying individual SNP effects between AD and cancer, we identified targets of the mentioned miRNAs. Changes in miRNA binding site due to SNPs can be compensated by redundancy in miRNA targeting (i.e. another miRNA can target site as a result of base change), which depends on codon degeneracy [95]. For instance, some members of miR-17 and miR-515 family have similar seed sequences – "AAGUGC" [96].

SNPs and miRNAs have immense potential in serving as genetic and blood-based biomarkers for diagnostic purposes, and further understanding the role of genetic predisposition will require studies of both cis and trans SNP effects. The genetic risk factors and associated miRNAs identified here would ideally be validated in a cross-conditional cohort of AD and cancer using microarray/RNA-seq to identify functional consequences of the SNPs implicated here which exert

cross-phenotypic effects between AD and cancer. Overall, these miRNAs play contrary roles in both diseases, making it imperative to investigate the strength of these miRNAs and identified targets to observe the extent of their influence in rescuing cognitive dysfunction in Alzheimer's disease. Our future work will investigate directionality of gene expression in AD & cancer under the influence of aggregated SNPs.

4.4.5 Limitations and considerations

We attempted to restrict phenotypic and technical bias for the investigation of this complex relationship between AD and cancer, however, the following considerations are important when interpreting these results. First, due to age and sex-stratification, our study is underpowered to detect all SNPs which may be exhibiting cross-effects between the two phenotypes. Second, since the odds ratios reported by epidemiological findings varies substantially (even for the same cancer types), we chose to remain conservative in our selection of priors for the Bayesian approach. Third, we were not able to use APOE as a covariate in the discovery phase. The genotyping of APOE is typically conducted independent of genome-wide SNP typing for AD studies, however genotyping of APOE is not typical of cancer genetic studies. To attempt to obtain APOE genotypes for the cancer cohorts, we relied on imputation, which wasn't successful for all individuals. As an alternative, we adjusted for *APOE* SNP (rs429358) as conditional test on merged populations for females and males in each TOMM40 haploblock to test for independent effects between *TOMM40* and *APOE.* Additionally, conditioning enabled detection of secondary associations (if any) in a local region which may have been otherwise obscured by the APOE effect. Fourth, factors such depression, education level and medical history are important consideration towards both AD and cancer. These variables were absent from the ADGC and BPC3 datasets obtained from dbGaP, therefore, such variables remained unaccounted. Since cancer history was available for ADNI

cohort, these individuals were removed from analysis; however, this information was not available for ADGC cohort. In order to adjust for any variable, it needs to be present for all the individuals, otherwise the individual is treated as missing. Finally, when we carried out analysis by employing a relaxed threshold on age (50-90) and higher priors, we get more hits including the ones being reported here, which we presume to be false positives or age-associated and not phenotype specific. This indicates that the analysis is sensitive to cohort selection procedures and we chose a more conservative approach to mitigate these confounding factors by employing a two-stage design with careful inclusion/exclusion criteria.

ACKNOWLEDGEMENTS

We like to acknowledge the NIH – Neurobiology of Aging T32 grant AG020494 for supporting this research. We also appreciate the datasets received via authorized access from ADNI, ADGC and BPC3.

The Breast and Prostate Cancer Cohort Consortium (BPC3) genome-wide association studies of advanced prostate cancer and estrogen-receptor negative breast cancer was supported by the National Cancer Institute under cooperative agreements U01-CA98233, U01-CA98710, U01-CA98216, and U01-CA98758 and the Intramural Research Program of the National Cancer Institute, Division of Cancer Epidemiology and Genetics.

Genotyping is performed by Alzheimer's Disease Genetics Consortium (ADGC), U01 AG032984, RC2AG036528. Phenotypic collection is coordinated by the National Alzheimer's Coordinating Center (NACC), U01AG016976. Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24

AG21886) awarded by the National Institute on Aging (NIA), were used in this study. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this study possible. Data for this study were prepared, archived, and distributed by the NIA Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689-01).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern

California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This project was supported in part, by the Texas Alzheimer's Research and Care Consortium by the Darrell K Royal Texas Alzheimer's Initiative, directed by the Texas Council on Alzheimer's Disease and Related Disorders.

COMPETING INTERESTS

Authors have no conflicts of interest to declare.

Table 1: *miRNAs expression in Alzheimer's disease and cancer. (References in supplementary file3 Table S1)*

miRNA	Cancer	Tissue [Ref]	Alzheimer's disease	Tissue [Ref]
hsa-miR-335	Downregulated	Serum [1] Plasma [2] Tumor [3]	Upregulated	Serum [4] Aged astrocytes & hippocampal brain [5]
hsa-miR-197	Upregulated	Tumor [6]	Downregulated	Serum [7]
	Downregulated	Tumor [8, 9]	Upregulated	Cortex, CSF [10]
hsa-miR-125b	Upregulated	Tumor [11, 12] , Serum [13], plasma [14]	Downregulated	Serum [15, 16]
	Downregulated	Tumor [17, 18]	Upregulated	CSF [19], frontal cortex [20]
hsa-miR-26b	Downregulated	Tumor [21, 22] Serum [23]	Upregulated	Serum [24]
hsa-miR-17	Upregulated	Serum [25]	Downregulated	Blood [26]
hsa-miR-20a	Upregulated	Tumor [27]	Downregulated	Brain [28]
hsa-miR-20b	Downregulated	Tumor [29, 30]	Upregulated	Serum [31]
hsa-miR-106b	Upregulated	Tumor [32]	Downregulated	Cortex [33]
	Downregulated	Serum [34]	Upregulated	Serum [35]
hsa-miR-93	Upregulated	Tumor [32, 36], Serum [36]	Downregulated	Serum [37]
hsa-miR-515	Upregulated	Tumor cell line[38]	Downregulated	Brain [39]
hsa-miR-34a	Downregulated	Tumor [40]	Upregulated	Brain [41]
hsa-miR-143	Downregulated	Tumor, cell line [42]	Upregulated	Brain [43]

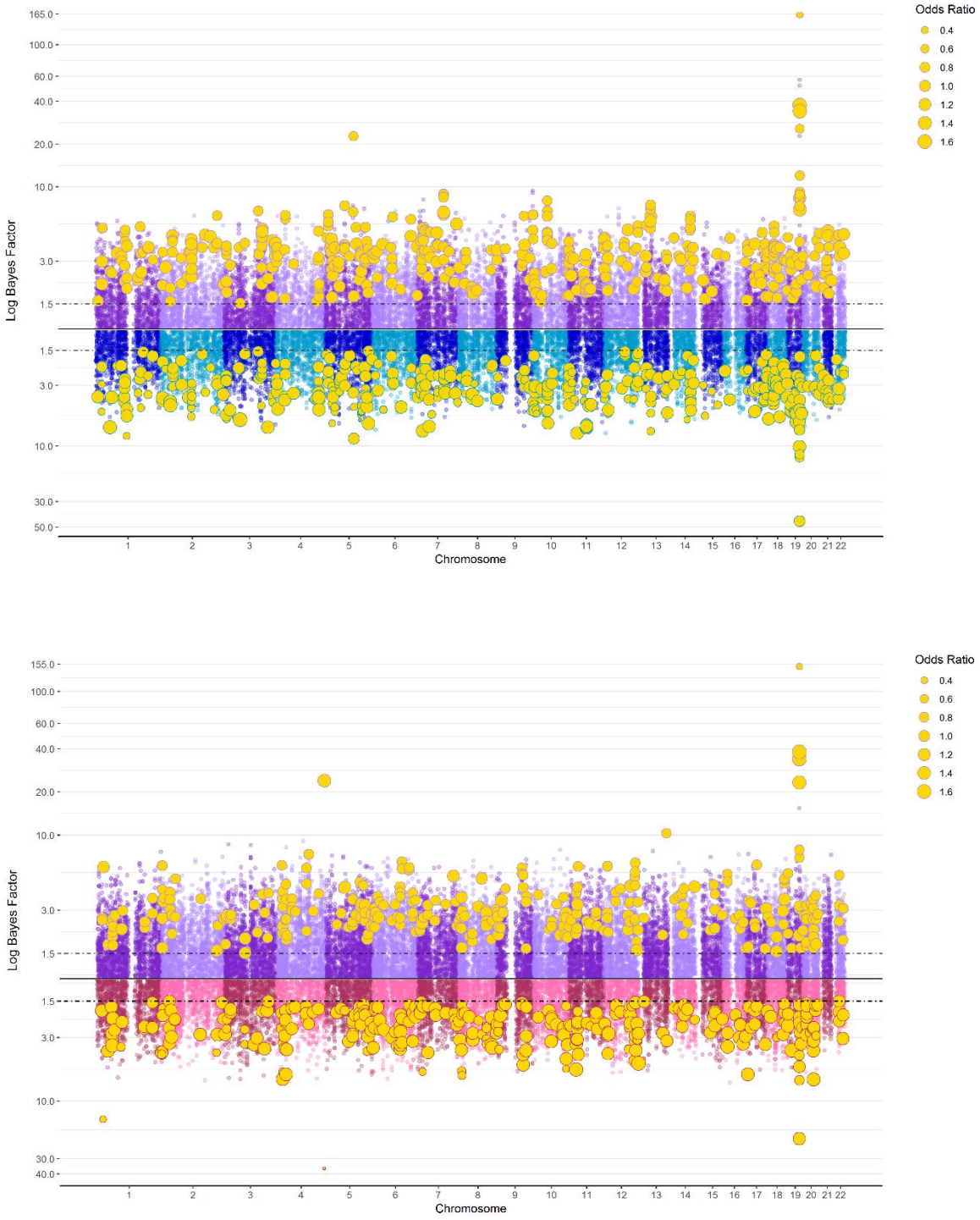


Figure 1. Discovery phase B-GWAS results. *Top panel: Bokeh Plot of multinomial GWAS comparing Alzheimer's disease and prostate cancer in the discovery stage. The purple Manhattan plot shows the results from Alzheimer's disease vs control, and the blue Manhattan plot shows the results of prostate cancer vs controls. The y axis is the logBF (log Bayes factor) for the respective disease, the significant SNPs are highlighted in yellow and their size is relative to the odds ratio as seen in the legend.*
Bottom panel: Bokeh Plot of multinomial GWAS comparing Alzheimer's disease and breast cancer in the discovery stage. The purple Manhattan plot shows the results from Alzheimer's disease vs control, and the pink Manhattan plot shows the results of

breast cancer vs controls. The y axis is the logBF (log Bayes factor) for the respective disease, the significant SNPs are highlighted in yellow and their size is relative to the odds ratio as seen in the legend.
[Note: Bokeh plots are an intersection between Manhattan and bubble plots]

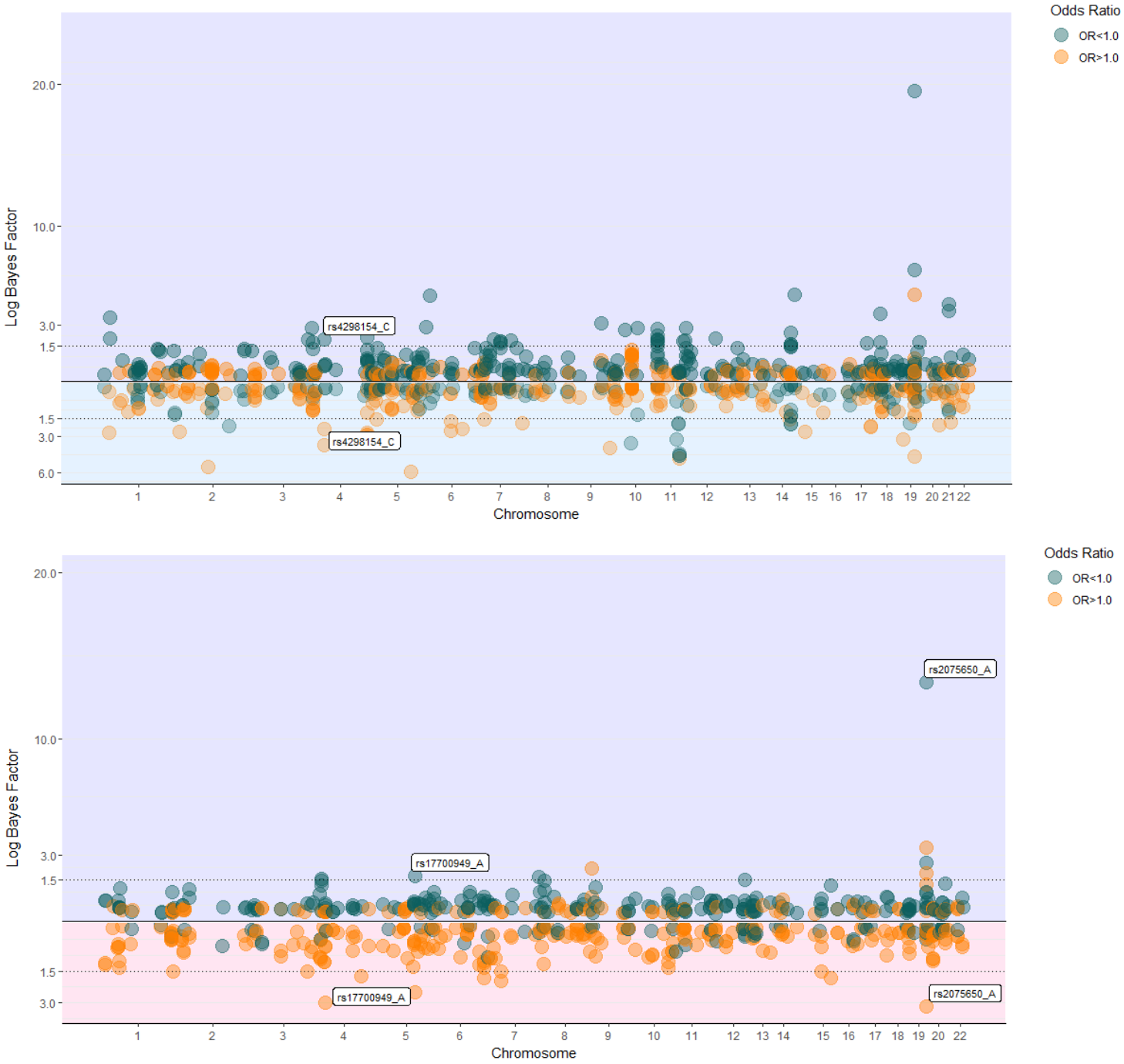


Figure 2. Replication stage B-GWAS results. *Top panel: Bokeh Plot of replication stage - multinomial GWAS comparing Alzheimer's disease and prostate cancer. Out of the 422 SNPs identified in the Discovery stage based on the criteria for inverse and significant (i.e., as odds ratio < 1 for Alzheimer's and odds ratio > 1 for cancer, or vice versa and an overall log Bayes factor of ≥ 3), 411 were typed in the Replication dataset. Organized by chromosome, the log Bayes factor for the analysis comparing AD to the common controls are shown in the upward facing Manhattan plot (shaded in light purple), and the log Bayes factor for the analysis comparing prostate cancer to the common controls are shown in the downward facing Manhattan plot (shaded in light*

blue). One SNP from the Discovery set were replicated based on inverse risk and strength of association, each indicated by the labeled data points: rs4298154.

Bottom panel: Bokeh Plot of replication stage - multinomial GWAS comparing Alzheimer's disease and breast cancer. Out of the 324 SNPs identified in the Discovery stage based on the criteria for inverse and significant (i.e., as odds ratio < 1 for Alzheimer's and odds ratio > 1 for cancer, or vice versa and an overall log Bayes factor of ≥ 3), 309 were typed in the Replication dataset. Organized by chromosome, the log Bayes factor for the analysis comparing AD to the common controls are shown in the upward facing Manhattan plot (shaded in light purple), and the log Bayes factor for the analysis comparing breast cancer to the common controls are shown in the downward facing Manhattan plot (shaded in light pink). Two SNPs from the Discovery set were replicated based on inverse risk and strength of association, each indicated by the labeled data points: rs17700949 and rs2075650.

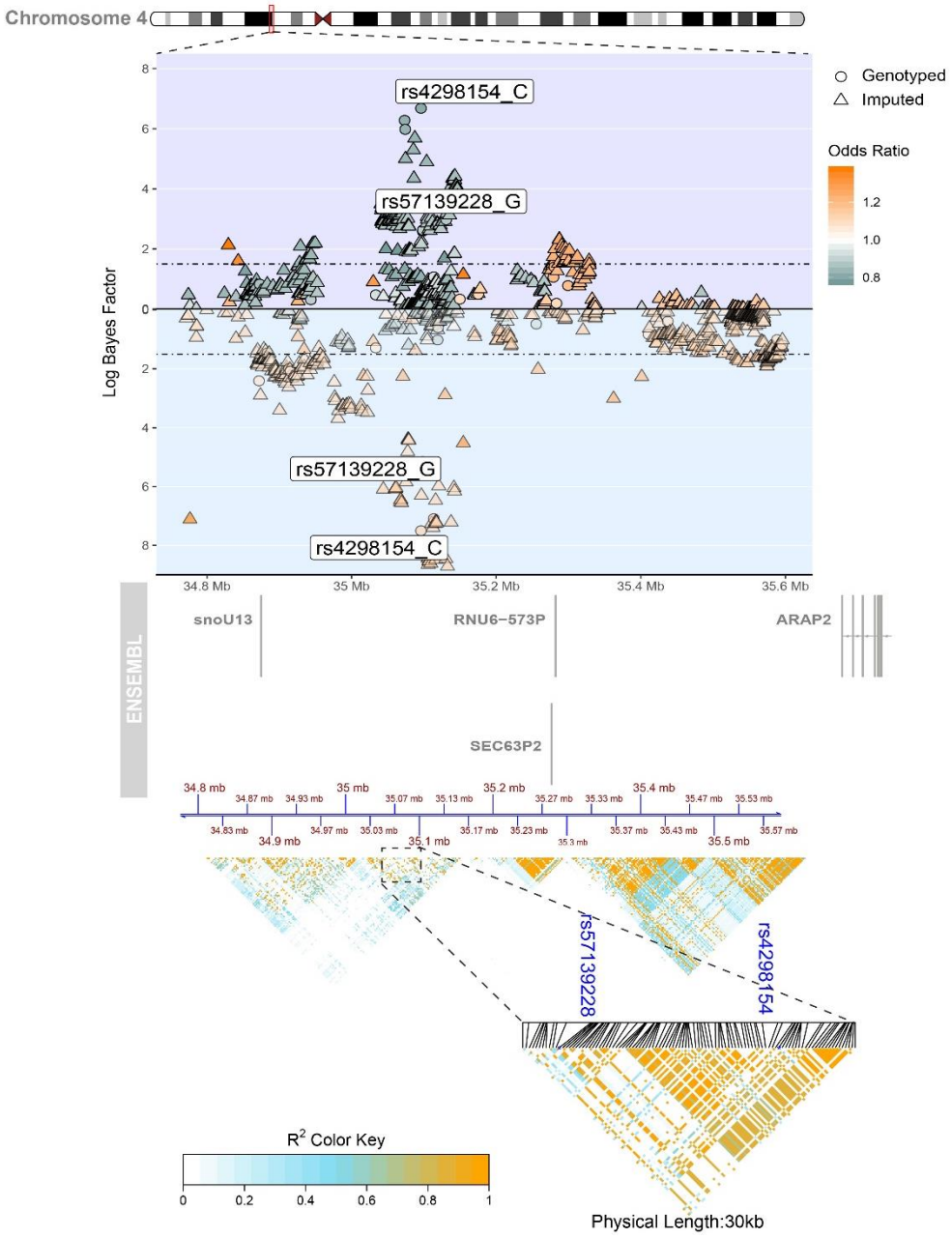


Figure 3: Regional Manhattan Plot *of Chromosome 4 risk region for association between males of Alzheimer's disease and prostate cancer. Left panel – In the Manhattan plot, the purple background shows the logBF (log Bayes factor) for Alzheimer's*

disease, and the blue background for prostate cancer. 23 SNPs were found to be significant in the 1Mbp - Chr4 risk region; the plot highlights most significant hits within the region. The genomic coordinates are shown using GrCh37 from Ensembl, followed my LD heat map of the corresponding region in the bottom panel. A zoomed-in LD map around the significant variants is shown at the bottom. Right panel – After conditioning on two SNPs with highest logBF, none of the variants were significant for both AD and cancer with opposite odds ratios.

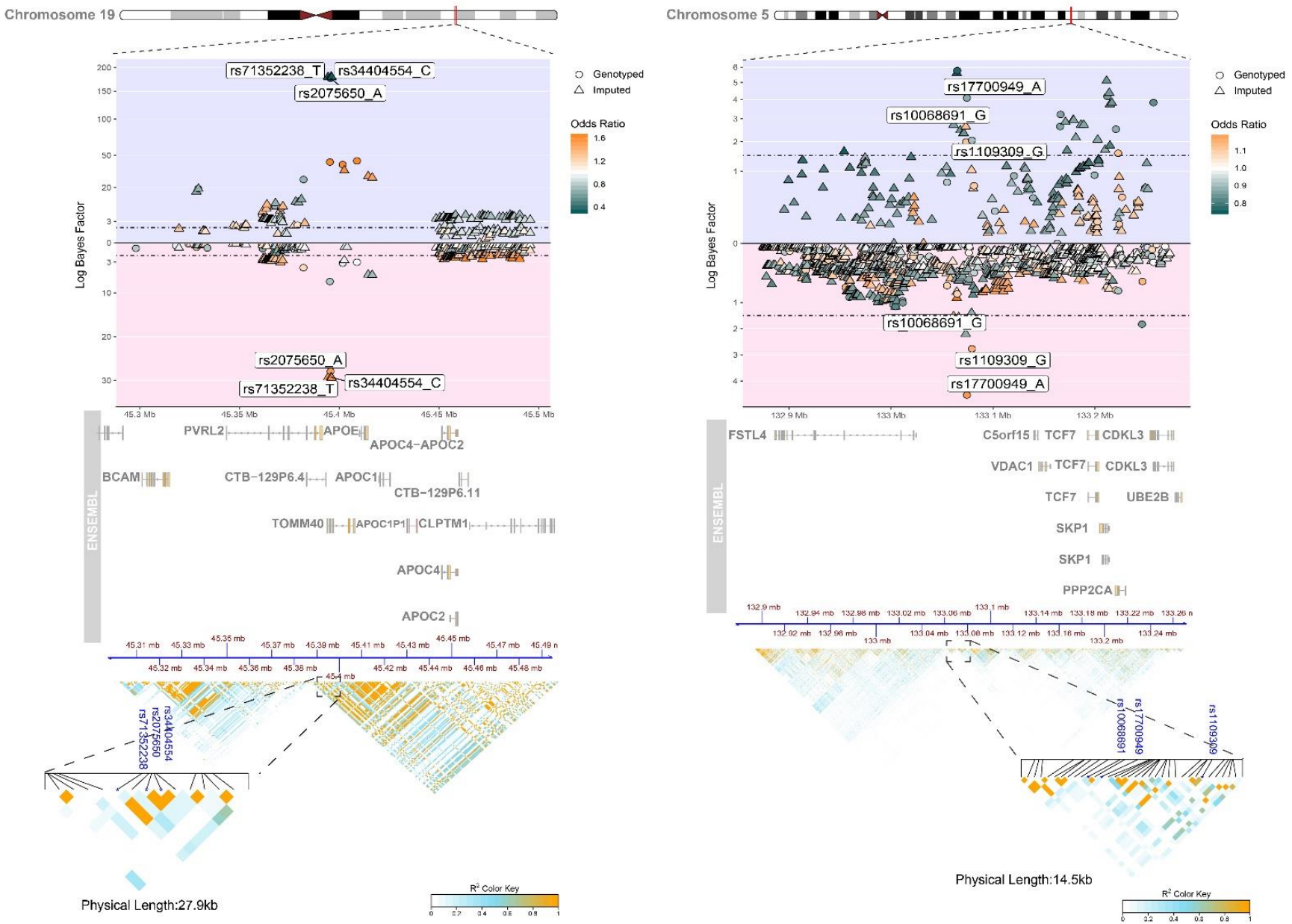


Figure 4: Regional Manhattan Plot. *Left panel. Chromosome 19 risk region for association between females of Alzheimer's and breast cancer. Left panel – In the Manhattan plot, the purple background shows the logBF (log Bayes factor) for Alzheimer's disease, and the pink background for breast cancer. 113 SNPs were found to be significant in the 1Mbp - Chr19 risk region; the*

plot highlights most significant hits within the region. The genomic coordinates are shown using GrCh37 from Ensembl, followed my LD heat map of the corresponding region in the bottom panel. A zoomed-in LD map around the significant variants is shown at the bottom.

Right panel Regional Manhattan Plot of Chromosome 5 risk region for association between females of Alzheimer's and breast cancer. In the Manhattan plot, the purple background shows the logBF (log Bayes factor) for Alzheimer's, and the pink background for breast cancer. 3 SNPs were found to be significant in the 1Mbp – Chr5 risk region and highlighted in the plot. The genomic coordinates are shown using GrCh37 from Ensembl, followed my LD heat map of the corresponding region in the bottom panel. A zoomed-in LD map around the significant variants is shown at the bottom.

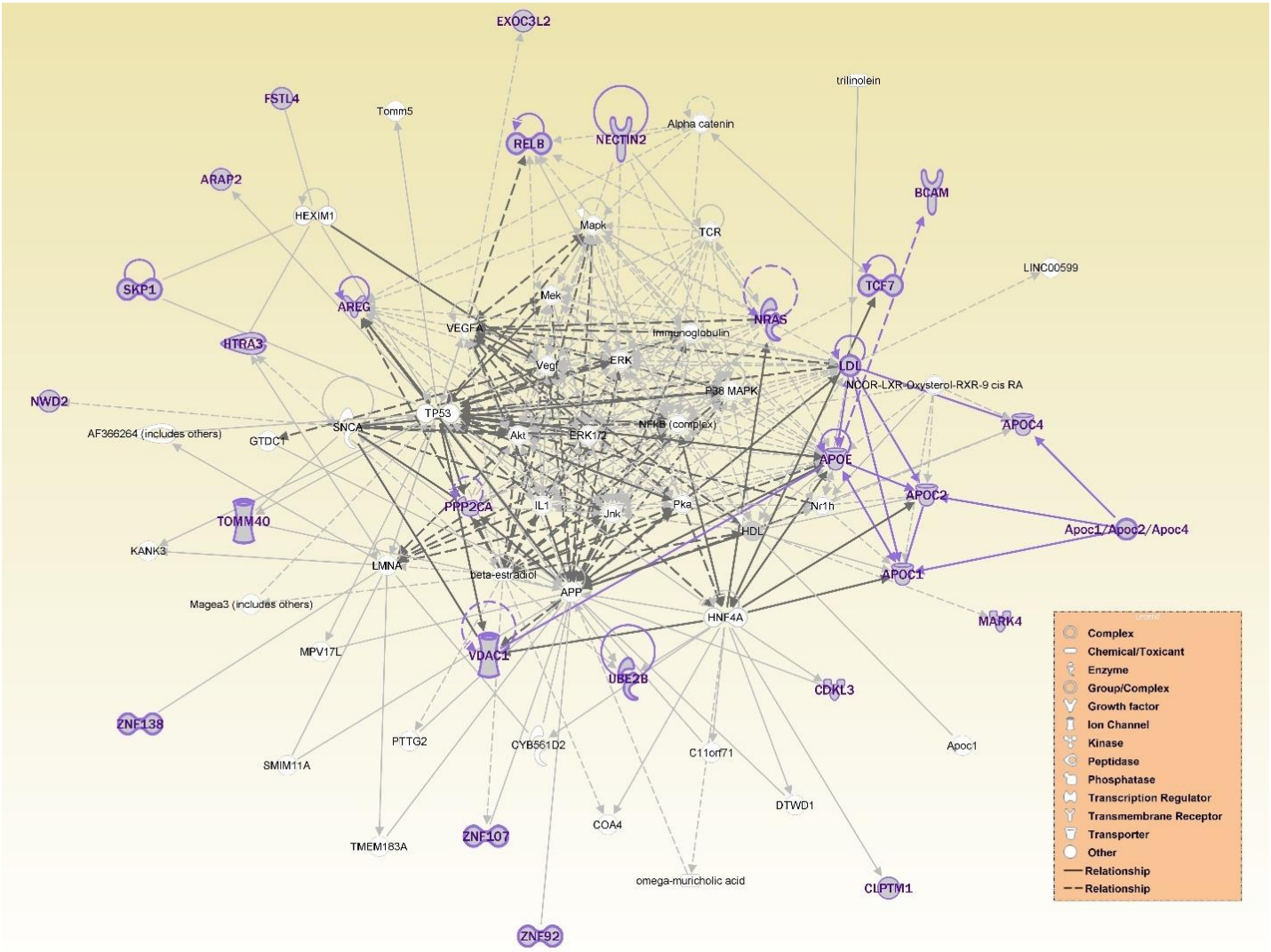


Figure 5: IPA network based on genes identified by B-GWAS. *Network created from query genes (highlighted in purple) using IPA's biobase knowledge.*

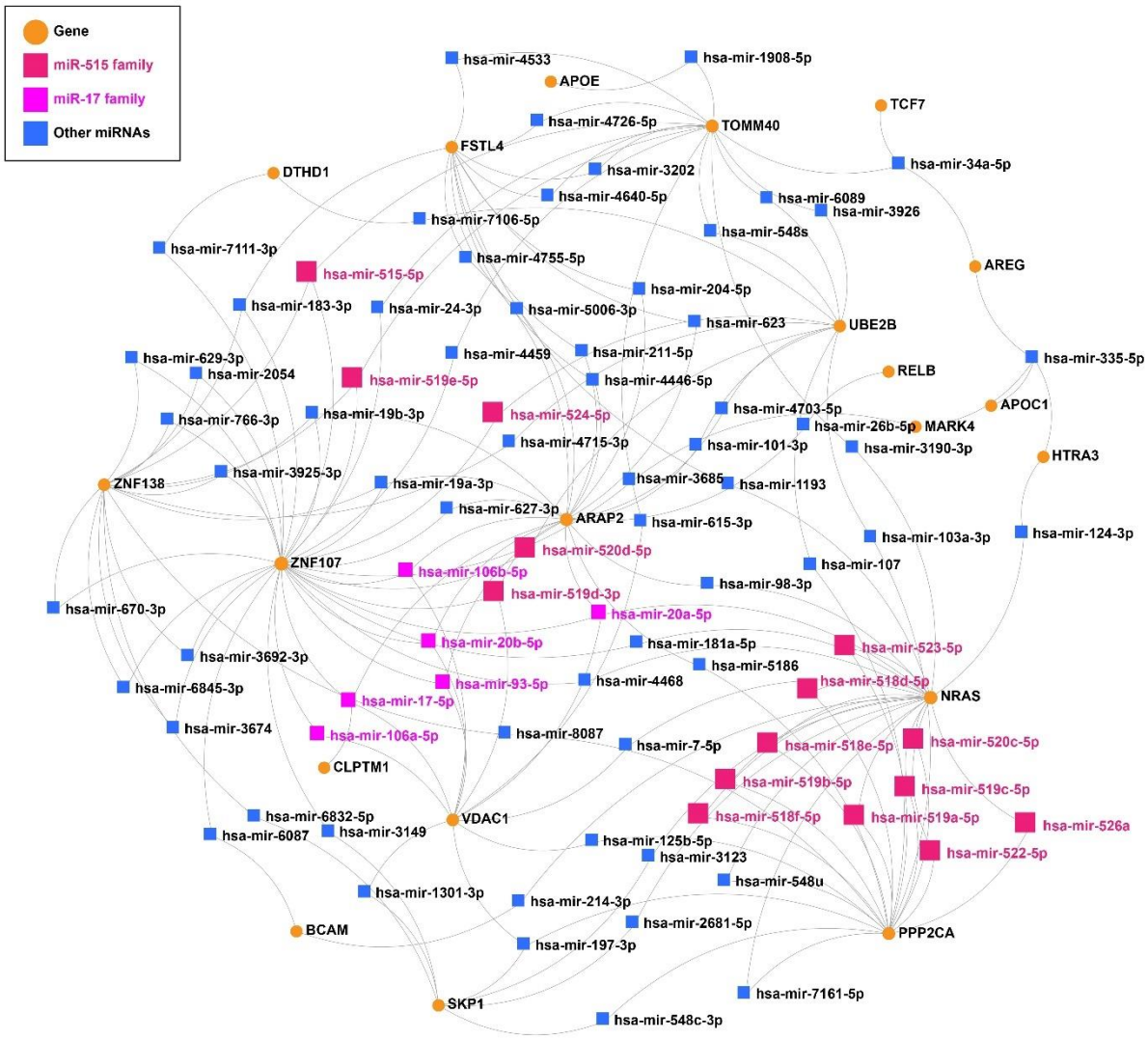


Figure 6: ***miRNA Network based on genes identified by B-GWAS****. The miRNA network was generated from a list of all implicated genes/loci identified in the B-GWAS; orange nodes are genes which are targeted by common miRNAs. Statistical enrichment of multiple family miRNAs was observed (dark blue squares), and several miRNAs exhibit high connectivity. Two miRNA families – miR-515 (dark pink squares) and miR-17 (magenta squares) show statistical enrichment.*

4.5 REFERENCES

[1] Mark Mather LAJ, Kelvin M. Pollard. Aging in the United States. Population Bulletin: Population Reference Bureau; 2015.
[2] Johnson SC, Dong X, Vijg J, Suh Y. Genetic evidence for common pathways in human age-related diseases. Aging Cell. 2015;14:809-17.
[3] Tabares-Seisdedos R, Rubenstein JL. Inverse cancer comorbidity: a serendipitous opportunity to gain insight into CNS disorders. Nat Rev Neurosci. 2013;14:293-304.
[4] Ma LL, Yu JT, Wang HF, Meng XF, Tan CC, Wang C, et al. Association between cancer and Alzheimer's disease: systematic review and meta-analysis. J Alzheimers Dis. 2014;42:565-73.
[5] Frain L, Swanson D, Cho K, Gagnon D, Lu KP, Betensky RA, et al. Association of cancer and Alzheimer's disease risk in a national cohort of veterans. Alzheimers Dement. 2017;13:1364-70.
[6] Ridge PG, Mukherjee S, Crane PK, Kauwe JSK, Alzheimer's Disease Genetics C. Alzheimer's Disease: Analyzing the Missing Heritability. PLOS ONE. 2013;8:e79771.
[7] Lilyquist J, Ruddy KJ, Vachon CM, Couch FJ. Common Genetic Variation and Breast Cancer Risk—Past, Present, and Future. 2018;27:380-94.
[8] Blanco-Gomez A, Castillo-Lluva S, Del Mar Saez-Freire M, Hontecillas-Prieto L, Mao JH, Castellanos-Martin A, et al. Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes. Bioessays. 2016;38:664-73.
[9] Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Human molecular genetics. 2015;24:R111-R9.
[10] Majumdar A, Haldar T, Bhattacharya S, Witte JS. An efficient Bayesian meta-analysis approach for studying cross-phenotype genetic associations. PLoS Genet. 2018;14:e1007139.
[11] Jostins L, McVean G. Trinculo: Bayesian and frequentist multinomial logistic regression for genome-wide association studies of multi-category phenotypes. Bioinformatics. 2016;32:1898-900.
[12] Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat. 2008;2:1360-83.
[13] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. 2017;67:7-30.
[14] 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.
[15] Hunter DJ, Riboli E, Haiman CA, Albanes D, Altshuler D, Chanock SJ, et al. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. Nat Rev Cancer. 2005;5:977-85.
[16] 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.
[17] Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010;5:1564-73.
[18] Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326-8.
[19] Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome research. 2009;19:1655-64.
[20] Wakefield J. Reporting and interpretation in genome-wide association studies. International Journal of Epidemiology. 2008;37:641-53.

[21] Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol. 2009;33:79-86.
[22] Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLOS Genetics. 2009;5:e1000529.
[23] Dayem Ullah AZ, Oscanoa J, Chelala C, Nagano A, Wang J, Lemoine NR. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Research. 2018;46:W109-W13.
[24] Ribeiro P, Arora SK, Fan Y, Xia J, Siklenka K, Kimmins S. miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. Nucleic Acids Research. 2016;44:W135-W41.
[25] Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: Triad of risk of Alzheimer's disease. The Journal of steroid biochemistry and molecular biology. 2016;160:134-47.
[26] Anand R, Prakash SS, Veeramanikandan R, Kirubakaran R. Association between apolipoprotein E genotype and cancer susceptibility: a meta-analysis. J Cancer Res Clin Oncol. 2014;140:1075-85.
[27] Roses A, Sundseth S, Saunders A, Gottschalk W, Burns D, Lutz M. Understanding the genetics of APOE and TOMM40 and role of mitochondrial structure and function in clinical pharmacology of Alzheimer's disease. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2016;12:687-94.
[28] Kim S, Cho H, Yang W, Kwon H, yeon Shin H, Lee E, et al. Abstract 2456: Overexpression of TOM40 (translocase in the outer mitochondrial membrane 40) inhibits the cell proliferation, invasion and migration abilities in ovarian cancer cell lines2014.
[29] Ning L, Pan B, Zhao Y-p, Liao Q, Zhang T-p, Chen G, et al. Immuno-proteomic screening of human pancreatic cancer associated membrane antigens for early diagnosis2007.
[30] Yashin AI, Fang F, Kovtun M, Wu D, Duan M, Arbeev K, et al. Hidden heterogeneity in Alzheimer's disease: Insights from genetic association studies and other analyses. Experimental Gerontology. 2018;107:148-60.
[31] Annadurai N, Agrawal K, Džubák P, Hajdúch M, Das VJC, Sciences ML. Microtubule affinity-regulating kinases are potential druggable targets for Alzheimer's disease. 2017;74:4159-69.
[32] Jenardhanan P, Mannu J, Mathur PP. The structural analysis of MARK4 and the exploration of specific inhibitors for the MARK family: a computational approach to obstruct the role of MARK4 in prostate cancer progression. Mol Biosyst. 2014;10:1845-68.
[33] Heidary Arash E, Shiban A, Song S, Attisano L. MARK4 inhibits Hippo signaling to promote proliferation and migration of breast cancer cells. EMBO reports. 2017;18:420-36.
[34] Hao S, Wang R, Zhang Y, Zhan H. Prediction of Alzheimer's Disease-Associated Genes by Integration of GWAS Summary Data and Expression Data. Frontiers in Genetics. 2019;9.
[35] Inoue K, Hatano K, Hanamatsu Y, Saigo C, Kito Y, Bunai K, et al. Pathobiological role of cleft palate transmembrane protein 1 family proteins in oral squamous cell carcinoma. 2019.
[36] Yang YC, Fu WP, Zhang J, Zhong L, Cai SX, Sun C. rs401681 and rs402710 confer lung cancer susceptibility by regulating TERT expression instead of CLPTM1L in East Asian populations. Carcinogenesis. 2018;39:1216-21.
[37] Koike Folgueira MAA, Carraro DM, Brentani H, da Costa Patrão DF, Barbosa EM, Netto MM, et al. Gene Expression Profile Associated with Response to Doxorubicin-Based Therapy in Breast Cancer. 2005;11:7434-43.

[38] Cudaback E, Li X, Yang Y, Yoo T, Montine KS, Craft S, et al. Apolipoprotein C-I is an APOE genotype-dependent suppressor of glial activation. 2012;9:192.
[39] Ren H, Chen Z, Yang L, Xiong W, Yang H, Xu K, et al. Apolipoprotein C1 (APOC1) promotes tumor progression via MAPK signaling pathways in colorectal cancer. Cancer management and research. 2019;11:4917-30.
[40] Giopanou I, Lilis I, Papadaki H, Papadas T, Stathopoulos GT. A link between RelB expression and tumor progression in laryngeal cancer. Oncotarget. 2017;8:114019-30.
[41] Nho K, Kim S, Horgusluoglu E, Risacher SL, Shen L, Kim D, et al. Association analysis of rare variants near the APOE region with CSF and neuroimaging biomarkers of Alzheimer's disease. 2017;10:29.
[42] Divers J, Palmer ND, Langefeld CD, Brown WM, Lu L, Hicks PJ, et al. Genome-wide association study of coronary artery calcified atherosclerotic plaque in African Americans with type 2 diabetes. BMC genetics. 2017;18:105.
[43] Melen E, Granell R, Kogevinas M, Strachan D, Gonzalez JR, Wjst M, et al. Genome-wide association study of body mass index in 23 000 individuals with and without asthma. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2013;43:463-74.
[44] Liu Y, Zhu D, Xing H, Hou Y, Sun Y. A 6-gene risk score system constructed for predicting the clinical prognosis of pancreatic adenocarcinoma patients. Oncology reports. 2019;41:1521-30.
[45] Lin VC, Huang C-Y, Lee Y-C, Yu C-C, Chang T-Y, Lu T-L, et al. Genetic variations in TP53 binding sites are predictors of clinical outcomes in prostate cancer patients. Archives of Toxicology. 2014;88:901-11.
[46] Boutahar N, Wierinckx A, Camdessanche JP, Antoine J-C, Reynaud E, Lassabliere F, et al. Differential effect of oxidative or excitotoxic stress on the transcriptional profile of amyotrophic lateral sclerosis-linked mutant SOD1 cultured neurons. 2011;89:1439-50.
[47] Shoshan-Barmatz V, Nahon-Crystal E, Shteinfer-Kuzmine A, Gupta R. VDAC1, mitochondrial dysfunction, and Alzheimer's disease. Pharmacological research. 2018;131:87-101.
[48] Mazure NM. VDAC in cancer. Biochimica et Biophysica Acta (BBA) - Bioenergetics. 2017;1858:665-73.
[49] McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nature Genetics. 2017;49:1126.
[50] Liu C, Chyr J, Zhao W, Xu Y, Ji Z, Tan H, et al. Genome-Wide Association and Mechanistic Studies Indicate That Immune Response Contributes to Alzheimer's Disease Development. Front Genet. 2018;9:410.
[51] Warrier V, Grasby KL, Uzefovsky F, Toro R, Smith P, Chakrabarti B, et al. Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and cognition. Molecular Psychiatry. 2017;23:1402.
[52] Guo Y, Tomlinson B, Chu T, Fang YJ, Gui H, Tang CS, et al. A Genome-Wide Linkage and Association Scan Reveals Novel Loci for Hypertension and Blood Pressure Traits. PLOS ONE. 2012;7:e31489.
[53] Ibanez K, Boullosa C, Tabares-Seisdedos R, Baudot A, Valencia A. Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. PLoS Genet. 2014;10:e1004173.
[54] Nudelman KNH, McDonald BC, Lahiri DK, Saykin AJJMN. Biological Hallmarks of Cancer in Alzheimer's Disease. 2019.

[55] Wang S, Wen P, Wood S. Effect of LXR/RXR agonism on brain and CSF A?40 levels in rats [version 2; peer review: 1 approved, 2 approved with reservations]. 2016;5.
[56] Ju X, Huang P, Chen M, Wang Q. Liver X receptors as potential targets for cancer therapeutics. Oncology letters. 2017;14:7676-80.
[57] Jęśko H, Wencel P, Strosznajder RP, Strosznajder JB. Sirtuins and Their Roles in Brain Aging and Neurodegenerative Disorders. Neurochem Res. 2017;42:876-90.
[58] Grasso M, Piscopo P, Confaloni A, Denti MA. Circulating miRNAs as biomarkers for neurodegenerative disorders. Molecules (Basel, Switzerland). 2014;19:6891-910.
[59] Nagaraj S, Zoltowska KM, Laskowska-Kaszub K, Wojda U. microRNA diagnostic panel for Alzheimer's disease and epigenetic trade-off between neurodegeneration and cancer. Aging Research Reviews. 2019;49:125-43.
[60] Basavaraju M, de Lencastre A. Alzheimer's disease: presence and role of microRNAs. Biomolecular concepts. 2016;7:241-52.
[61] Zhao J, Yue D, Zhou Y, Jia L, Wang H, Guo M, et al. The Role of MicroRNAs in Aβ Deposition and Tau Phosphorylation in Alzheimer's Disease. Frontiers in neurology. 2017;8:342-.
[62] Li H, Wu Q, Li T, Liu C, Xue L, Ding J, et al. The miR-17-92 cluster as a potential biomarker for the early diagnosis of gastric cancer: evidence and literature review. Oncotarget. 2017;8:45060-71.
[63] Pardo OE, Castellano L, Munro CE, Hu Y, Mauri F, Krell J, et al. miR-515-5p controls cancer cell migration through MARK4 regulation. EMBO reports. 2016;17:570-84.
[64] Gorospe M, Abdelmohsen K. MicroRegulators come of age in senescence. Trends in genetics : TIG. 2011;27:233-41.
[65] Pichler S, Gu W, Hartl D, Gasparoni G, Leidinger P, Keller A, et al. The miRNome of Alzheimer's disease: consistent downregulation of the miR-132/212 cluster. Neurobiology of Aging. 2017;50:167.e1-.e10.
[66] Holohan KN, Lahiri DK, Schneider BP, Foroud T, Saykin AJ. Functional microRNAs in Alzheimer's disease and cancer: differential regulation of common mechanisms and pathways. Front Genet. 2013;3.
[67] Wang M, Qin L, Tang B. MicroRNAs in Alzheimer's Disease. Frontiers in genetics. 2019;10:153-.
[68] Wang H, Tan G, Dong L, Cheng L, Li K, Wang Z, et al. Circulating MiR-125b as a marker predicting chemoresistance in breast cancer. PloS one. 2012;7:e34210-e.
[69] Sandoval-Bórquez A, Polakovicova I, Carrasco-Véliz N, Lobos-González L, Riquelme I, Carrasco-Avino G, et al. MicroRNA-335-5p is a potential suppressor of metastasis and invasion in gastric cancer. Clin Epigenetics. 2017;9:114-.
[70] Tomé M, Sepúlveda JC, Delgado M, Andrades JA, Campisi J, González MA, et al. miR-335 correlates with senescence/aging in human mesenchymal stem cells and inhibits their therapeutic actions through inhibition of AP-1 activity. Stem Cells. 2014;32:2229-44.
[71] Szybińska A, Leśniak W. P53 Dysfunction in Neurodegenerative Diseases - The Cause or Effect of Pathological Changes? Aging and disease. 2017;8:506-18.
[72] Raihan O, Brishti A, Molla MR, Li W, Zhang Q, Xu P, et al. The Age-dependent Elevation of miR-335-3p Leads to Reduced Cholesterol and Impaired Memory in Brain. Neuroscience. 2018;390:160-73.
[73] Absalon S, Kochanek DM, Raghavan V, Krichevsky AM. MiR-26b, upregulated in Alzheimer's disease, activates cell cycle entry, tau-phosphorylation, and apoptosis in postmitotic

neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2013;33:14645-59.
[74] Li Y, Sun Z, Liu B, Shan Y, Zhao L, Jia L. Tumor-suppressive miR-26a and miR-26b inhibit cell aggressiveness by regulating FUT4 in colorectal cancer. Cell death & disease. 2017;8:e2892.
[75] Verghese ET, Drury R, Green CA, Holliday DL, Lu X, Nash C, et al. MiR-26b is down-regulated in carcinoma-associated fibroblasts from ER-positive breast cancers leading to enhanced cell migration and invasion. The Journal of pathology. 2013;231:388-99.
[76] Sarkar S, Jun S, Rellick S, Quintana DD, Cavendish JZ, Simpkins JW. Expression of microRNA-34a in Alzheimer's disease brain targets genes linked to synaptic plasticity, energy metabolism, and resting state network activity. Brain Res. 2016;1646:139-51.
[77] Okada N, Lin C-P, Ribeiro MC, Biton A, Lai G, He X, et al. A positive feedback between p53 and miR-34 miRNAs mediates tumor suppression. Genes Dev. 2014;28:438-50.
[78] Slabáková E, Culig Z, Remšík J, Souček K. Alternative mechanisms of miR-34a regulation in cancer. Cell death & disease. 2017;8:e3100-e.
[79] Xu Y, Chen P, Wang X, Yao J, Zhuang S. miR-34a deficiency in APP/PS1 mice promotes cognitive function by increasing synaptic plasticity via AMPA and NMDA receptors. Neurosci Lett. 2018;670:94-104.
[80] Moszyńska A, Gebert M, Collawn JF, Bartoszewski R. SNPs in microRNA target sites and their potential role in human disease. Open biology. 2017;7:170019.
[81] Zhao J, Chen Y, Liu F, Yin M. Overexpression of miRNA-143 Inhibits Colon Cancer Cell Proliferation by Inhibiting Glucose Uptake. Archives of medical research. 2018;49:497-503.
[82] Llorens F, Thune K, Andres-Benito P, Tahir W, Ansoleaga B, Hernandez-Ortega K, et al. MicroRNA Expression in the Locus Coeruleus, Entorhinal Cortex, and Hippocampus at Early and Middle Stages of Braak Neurofibrillary Tangle Pathology. Journal of molecular neuroscience : MN. 2017;63:206-15.
[83] Su WZ, Ren LF. MiRNA-199 inhibits malignant progression of lung cancer through mediating RGS17. European review for medical and pharmacological sciences. 2019;23:3390-400.
[84] Zhang Y, Wang Y, Wang J. MicroRNA-584 inhibits cell proliferation and invasion in non-small cell lung cancer by directly targeting MTDH. Exp Ther Med. 2018;15:2203-11.
[85] Momen-Heravi F, Trachtenberg AJ, Kuo WP, Cheng YS. Genomewide Study of Salivary MicroRNAs for Detection of Oral Cancer. Journal of dental research. 2014;93:86s-93s.
[86] Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic acids research. 2014;42:D86-D91.
[87] Wang K, Long B, Jiao J-Q, Wang J-X, Liu J-P, Li Q, et al. miR-484 regulates mitochondrial network through targeting Fis1. Nature Communications. 2012;3:781.
[88] Williams M, Caino MC. Mitochondrial Dynamics in Type 2 Diabetes and Cancer. Frontiers in endocrinology. 2018;9:211-.
[89] Wang X, Su B, Lee H-g, Li X, Perry G, Smith MA, et al. Impaired Balance of Mitochondrial Fission and Fusion in Alzheimer's Disease. 2009;29:9090-103.
[90] Wilk G, Braun R. regQTLs: Single nucleotide polymorphisms that modulate microRNA regulation of gene expression in tumors. PLOS Genetics. 2018;14:e1007837.
[91] Saunders MA, Liang H, Li W-H. Human polymorphism at microRNAs and microRNA target sites. Proceedings of the National Academy of Sciences of the United States of America. 2007;104:3300-5.

[92] Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2014;1842:1910-22.
[93] Takousis P, Sadlon A, Schulz J, Wohlers I, Dobricic V, Middleton L, et al. Differential expression of microRNAs in Alzheimer's disease brain, blood and cerebrospinal fluid: a systematic review and meta-analysis. 2018:499491.
[94] Rahman MR, Islam T, Zaman T, Shahjaman M, Karim MR, Holsinger D, et al. Blood-based molecular biomarker signatures in Alzheimer's disease: Insights from systems biomedicine perspective. 2018:481879.
[95] Maleszka R, Mason PH, Barron AB. Epigenomics and the concept of degeneracy in biological systems. Briefings in Functional Genomics. 2013;13:191-202.
[96] Malnou EC, Umlauf D, Mouysset M, Cavaillé J. Imprinted MicroRNA Gene Clusters in the Evolution, Development, and Functions of Mammalian Placenta. Frontiers in genetics. 2019;9:706-.

CAPTIONS OF SUPPLEMENTARY FILES

- **Supplementary file1**
  - Table S1. ***Discovery stage - Bayesian multinomial regression comparing AD males (Disease1) and prostate cancer males (Disease3) against control males.*** *The results shown here are significant SNPs defined as OR<1 for Disease1 and OR>1 for Disease3, or vice versa and overall logBF of 3+, and individual logBF (M_Disease1/M_Disease3) of 1.5+.*
  - Table S2. ***Replication stage - Bayesian multinomial regression comparing AD males (Disease1) and prostate cancer males (Disease3) against control males.*** *The results shown here are SNPs which were found significant in the discovery stage, the SNPs that were significant are highlighted in blue, hereafter referred to as replicated hits/SNPs.*
  - Table S3. ***Discovery stage - Bayesian multinomial regression comparing AD females (Disease1) and breast cancer females (Disease2) against control females.*** *The results shown here are significant SNPs defined as OR<1 for Disease1 and OR>1 for Disease2, or vice versa and overall logBF of 3+, and individual logBF (M_Disease1/M_Disease2) of 1.5+.*
  - Table S4. ***Replication stage - Bayesian multinomial regression comparing AD females (Disease1) and breast cancer males (Disease2) against control females.*** *The results shown here are SNPs which were found significant in the discovery stage, the SNPs that were significant are highlighted in pink, hereafter referred to as replicated hits/SNPs.*

- Table S5. ***Regional GWAS - Bayesian multinomial regression comparing AD males (Disease1) and prostate cancer males (Disease3) against control males in chromosome 4.*** *The results shown here are significant SNPs defined as OR<1 for Disease1 and OR>1 for Disease3, or vice versa and overall logBF of 3+, and individual logBF (M_Disease1/M_Disease3) of 1.5+.*
- Table S6. ***Regional GWAS- Bayesian multinomial regression comparing AD females (Disease1) and breast cancer females (Disease2) against control females in chromosome 19.*** *The results shown here are significant SNPs defined as OR<1 for Disease1 and OR>1 for Disease2, or vice versa and overall logBF of 3+, and individual logBF (M_Disease1/M_Disease2) of 1.5+.*
- Table S7. ***Regional GWAS- Bayesian multinomial regression comparing AD females (Disease1) and breast cancer females (Disease2) against control females in chromosome 5.*** *The results shown here are significant SNPs defined as OR<1 for Disease1 and OR>1 for Disease2, or vice versa and overall logBF of 3+, and individual logBF (M_Disease1/M_Disease2) of 1.5+.*
- Table S8. ***Results in same direction identified in Discovery stage.*** *The results shown here are significant SNPs defined as OR<1 for Disease1 and OR<1 for Disease2/3, or vice versa and overall logBF of 3+, and individual logBF (M_Disease1/M_Disease2/3) of 1.5+.*

Link - Supplementaryfile1-AdvsCa

- **Supplementary file2**
  - *Functional annotation of significant SNPs, each tab denotes the name of the functional scoring method.*

Link- Suplementaryfile2-ADvsCa

- **Supplementary file3**
  - *Text S1. Methods and Material*
  - *Figure S1. Visual summary of methodology*
  - *Figure S2: Details of merging & QC of datasets*
  - *Table S1: MiRNAs expression in Alzheimer's disease and cancer*
  - *Table S2. Enrichment results of the miRNA family in the network using miRNet*
  - *Table S3. Details of the miRNA network visualized in Figure6*
  - *Table S4. miRNA target sites associated with query SNP- rs6859*
  - *Table S5. miR target site associated with query SNP- rs11556505*
  - *Figure S3. Enriched processes in reported genes*
  - *Figure S4. Network analysis of reported genes*
  - *Figure S5. Potential mechanism of SNPs-miRNA binding sites*

4.6 SUPPLEMENTARY FILE3

4.6.1 Text S1. Methods and Material

Datasets

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org. The Breast and Prostate Cancer Cohort Consortium (BPC3) was established in 2003 to pool data and biospecimens from nine large prospective cohorts to conduct research on gene-environment interactions in cancer etiology. The BPC3 GWAS includes the following cohorts: the American Cancer Society Cancer Prevention Study-II (CPS-II); the European Prospective Investigation of Cancer (EPIC); the Physician's Health Study (PHS); the Nurses' Health Studies I and II (NHS and NHSII); the Health Professionals Follow-up Study (HPFS); the Multiethnic Cohort (MEC); the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial; and the Alpha-Tocopherol, Beta-Carotene (ATBC) Study (https://epi.grants.cancer.gov/BPC3/). The Alzheimer's Disease Genetics Consortium is an $18.3 million five-year research grant to conduct genome-wide association studies (GWAS) to identify genes associated with an increased risk of developing late-onset Alzheimer's disease (LOAD) (http://www.adgenetics.org/).

4.6.2 Figure S1. Visual summary of methodology

An abridged visualization of the overall methods involved in the study, and the name of each supplementary file which contains the corresponding results is shown under each section.

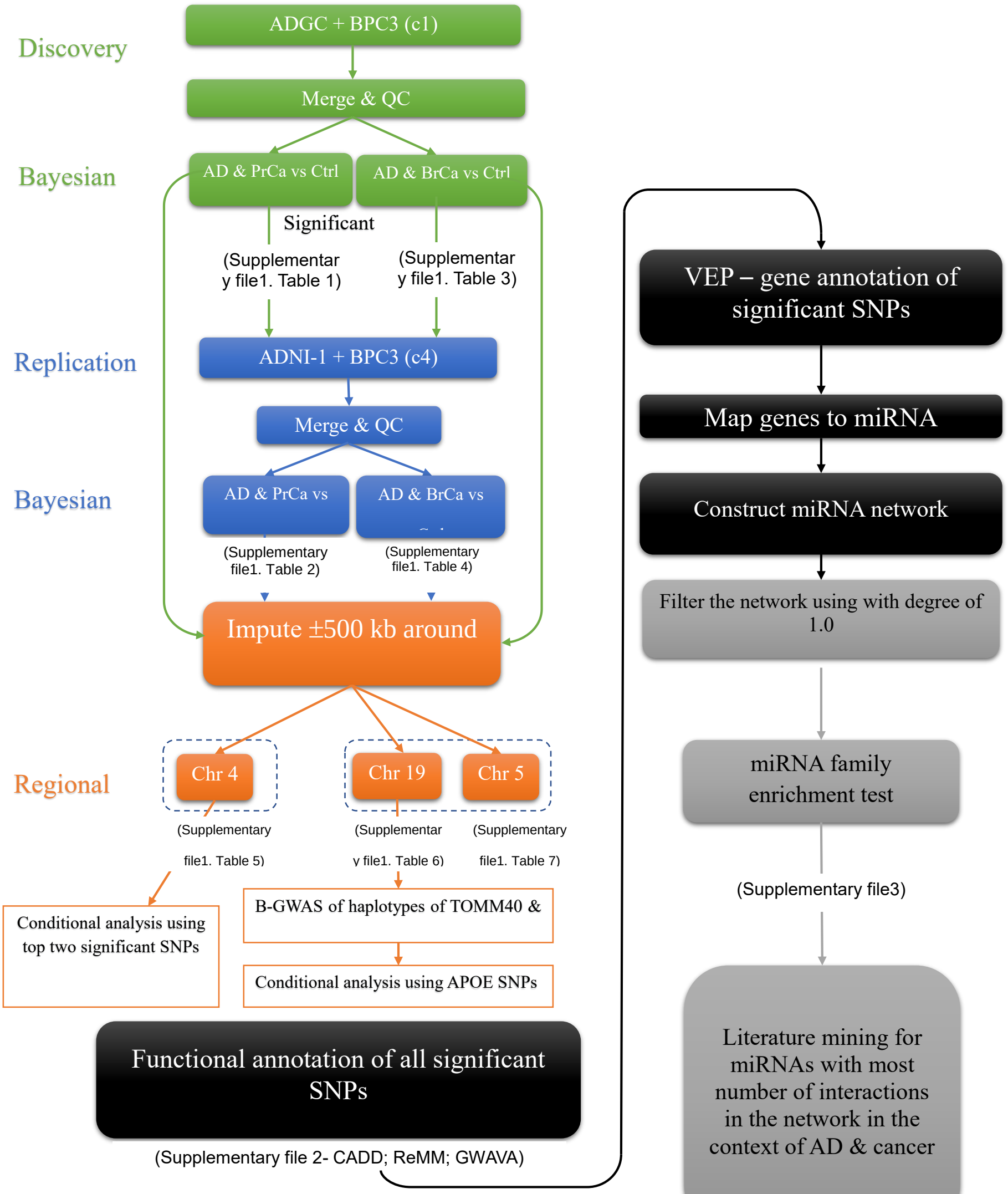
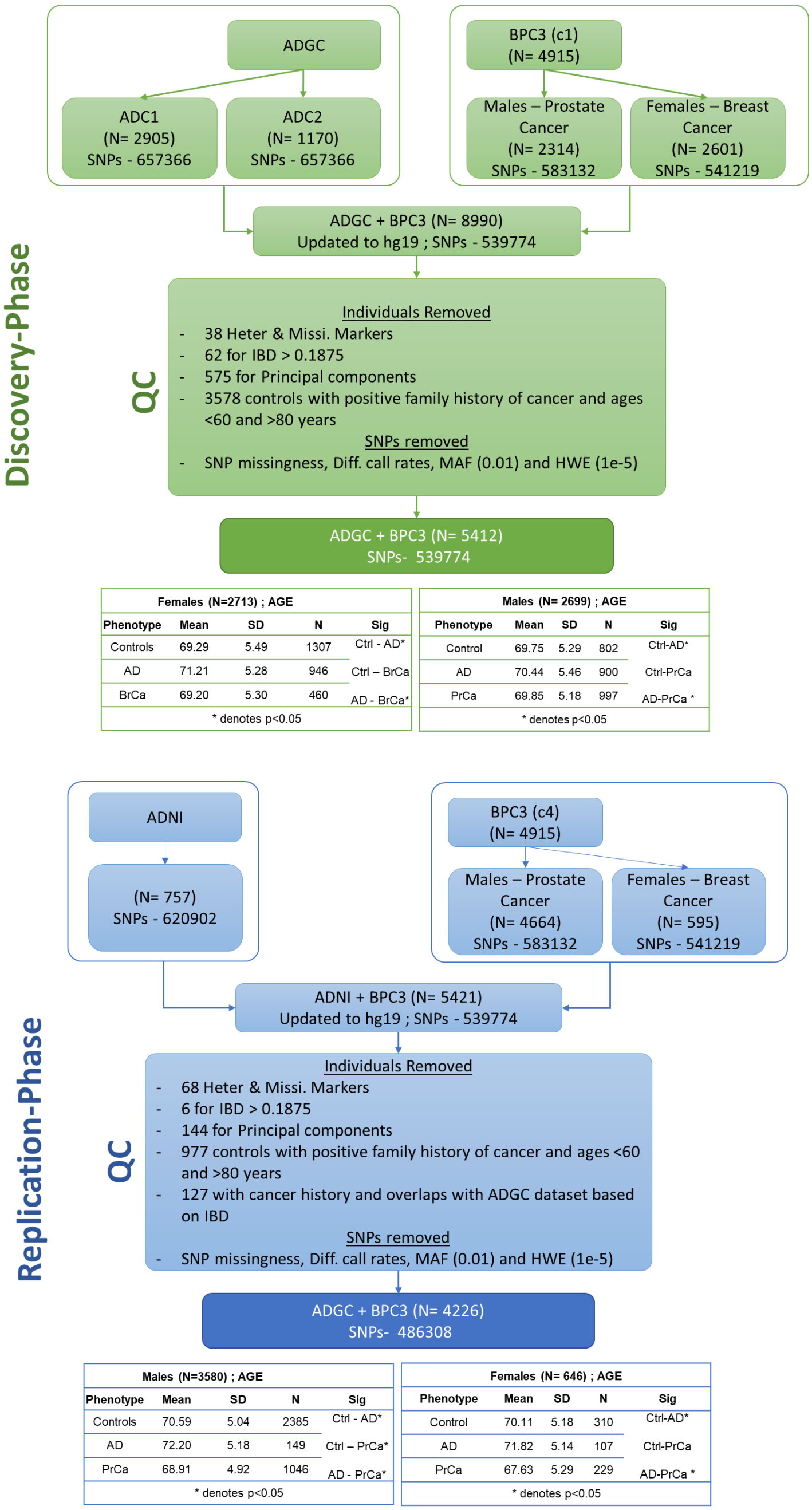


Figure S2: Details of merging & QC of datasets



Females (N=2713) ; AGE				
Phenotype	Mean	SD	N	Sig
Controls	69.29	5.49	1307	Ctrl - AD*
AD	71.21	5.28	946	Ctrl – BrCa
BrCa	69.20	5.30	460	AD - BrCa*
* denotes p<0.05				

Males (N= 2699) ; AGE				
Phenotype	Mean	SD	N	Sig
Control	69.75	5.29	802	Ctrl-AD*
AD	70.44	5.46	900	Ctrl-PrCa
PrCa	69.85	5.18	997	AD-PrCa *
* denotes p<0.05				

Males (N=3580) ; AGE				
Phenotype	Mean	SD	N	Sig
Controls	70.59	5.04	2385	Ctrl - AD*
AD	72.20	5.18	149	Ctrl – PrCa*
PrCa	68.91	4.92	1046	AD - PrCa*
* denotes p<0.05				

Females (N= 646) ; AGE				
Phenotype	Mean	SD	N	Sig
Control	70.11	5.18	310	Ctrl-AD*
AD	71.82	5.14	107	Ctrl-PrCa
PrCa	67.63	5.29	229	AD-PrCa *
* denotes p<0.05				

4.6.3 Table S1: MiRNAs expression in Alzheimer's disease and cancer.

miRNA	Cancer	Tissue [Ref]	Alzheimer's disease	Tissue [Ref]
hsa-miR-335	Downregulated	Serum [1] Plasma [2] Tumor [3]	Upregulated	Serum [4] Aged astrocytes & hippocampal brain [5]
hsa-miR-197	Upregulated	Tumor [6]	Downregulated	Serum [7]
	Downregulated	Tumor [8, 9]	Upregulated	Cortex, CSF [10]
hsa-miR-125b	Upregulated	Tumor [11, 12] , Serum [13], plasma [14]	Downregulated	Serum [15, 16]
	Downregulated	Tumor [17, 18]	Upregulated	CSF [19], frontal cortex [20]
hsa-miR-26b	Downregulated	Tumor [21, 22] Serum [23]	Upregulated	Serum [24]
hsa-miR-17	Upregulated	Serum [25]	Downregulated	Blood [26]
hsa-miR-20a	Upregulated	Tumor [27]	Downregulated	Brain [28]
hsa-miR-20b	Downregulated	Tumor [29, 30]	Upregulated	Serum [31]
hsa-miR-106b	Upregulated	Tumor [32]	Downregulated	Cortex [33]
	Downregulated	Serum [34]	Upregulated	Serum [35]
hsa-miR-93	Upregulated	Tumor [32, 36], Serum [36]	Downregulated	Serum [37]
hsa-miR-515	Upregulated	Tumor cell line[38]	Downregulated	Brain [39]
hsa-miR-34a	Downregulated	Tumor [40]	Upregulated	Brain [41]
hsa-miR-143	Downregulated	Tumor, cell line [42]	Upregulated	Brain [43]

References

[1] Cui L, Hu Y, Bai B, Zhang S. Serum miR-335 Level is Associated with the Treatment Response to Trans-Arterial Chemoembolization and Prognosis in Patients with Hepatocellular Carcinoma. Cellular Physiology and Biochemistry. 2015;37:276-83.
[2] Khalighfard S, Alizadeh AM, Irani S, Omranipour R. Plasma miR-21, miR-155, miR-10b, and Let-7a as the potential biomarkers for the monitoring of breast cancer patients. Scientific reports. 2018;8:17981-.
[3] Cao J, Cai J, Huang D, Han Q, Yang Q, Li T, et al. miR-335 represents an invasion suppressor gene in ovarian cancer by targeting Bcl-w. Oncology reports. 2013;30:701-6.
[4] Cheng L, Doecke JD, Sharples RA, Villemagne VL, Fowler CJ, Rembach A, et al. Prognostic serum miRNA biomarkers associated with Alzheimer's disease shows concordance with neuropsychological and neuroimaging assessment. Mol Psychiatry. 2015;20:1188-96.
[5] Raihan O, Brishti A, Molla MR, Li W, Zhang Q, Xu P, et al. The Age-dependent Elevation of miR-335-3p Leads to Reduced Cholesterol and Impaired Memory in Brain. Neuroscience. 2018;390:160-73.
[6] Lehmann U, Streichert T, Otto B, Albat C, Hasemeier B, Christgen H, et al. Identification of differentially expressed microRNAs in human male breast cancer. BMC cancer. 2010;10:109-.
[7] Jiang N, Ruan C-J, Cheng X-R, Wang L-N, Tan J-P, Wang W-S, et al. Meta-microRNAs as potential noninvasive markers for early diagnosis of Alzheimer's disease. bioRxiv. 2018:281915.
[8] Ni JS, Zheng H, Huang ZP, Hong YG, Ou YL, Tao YP, et al. MicroRNA-197-3p acts as a prognostic marker and inhibits cell invasion in hepatocellular carcinoma. Oncol Lett. 2019;17:2317-27.
[9] Mavridis K, Gueugnon F, Petit-Courty A, Courty Y, Barascu A, Guyetant S, et al. The oncomiR miR-197 is a novel prognostic indicator for non-small cell lung cancer patients. British Journal Of Cancer. 2015;112:1527.
[10] Maes OC, Chertkow HM, Wang E, Schipper HM. MicroRNA: Implications for Alzheimer Disease and other Human CNS Disorders. Current genomics. 2009;10:154-68.
[11] Luo Y, Wang X, Niu W, Wang H, Wen Q, Fan S, et al. Elevated microRNA-125b levels predict a worse prognosis in HER2-positive breast cancer patients. Oncology letters. 2017;13:867-74.
[12] Sui M, Jiao A, Zhai H, Wang Y, Wang Y, Sun D, et al. Upregulation of miR-125b is associated with poor prognosis and trastuzumab resistance in HER2-positive gastric cancer. Experimental and therapeutic medicine. 2017;14:657-63.
[13] Zuberi M, Khan I, Mir R, Gandhi G, Ray PC, Saxena A. Utility of Serum miR-125b as a Diagnostic and Prognostic Indicator and Its Alliance with a Panel of Tumor Suppressor Genes in Epithelial Ovarian Cancer. PLOS ONE. 2016;11:e0153902.
[14] Jiang Y, Luan Y, Chang H, Chen G. The diagnostic and prognostic value of plasma microRNA-125b-5p in patients with multiple myeloma. Oncol Lett. 2018;16:4001-7.
[15] Tan L, Yu JT, Liu QY, Tan MS, Zhang W, Hu N, et al. Circulating miR-125b as a biomarker of Alzheimer's disease. Journal of the neurological sciences. 2014;336:52-6.
[16] Galimberti D, Villa C, Fenoglio C, Serpente M, Ghezzi L, Cioffi SM, et al. Circulating miRNAs as potential biomarkers in Alzheimer's disease. J Alzheimers Dis. 2014;42:1261-7.
[17] Banzhaf-Strathmann J, Edbauer DJCC, Signaling. Good guy or bad guy: the opposing roles of microRNA 125b in cancer. 2014;12:30.
[18] Huang K, Dong S, Li W, Xie Z. The expression and regulation of microRNA-125b in cancers. Acta biochimica et biophysica Sinica. 2013;45:803-5.
[19] Alexandrov PN, Dua P, Hill JM, Bhattacharjee S, Zhao Y, Lukiw WJ. microRNA (miRNA) speciation in Alzheimer's disease (AD) cerebrospinal fluid (CSF) and extracellular fluid (ECF). International journal of biochemistry and molecular biology. 2012;3:365-73.
[20] Banzhaf-Strathmann J, Benito E, May S, Arzberger T, Tahirovic S, Kretzschmar H, et al. MicroRNA-125b induces tau hyperphosphorylation and cognitive deficits in Alzheimer's disease. The EMBO journal. 2014;33:1667-80.

[21] Cao J, Guo T, Dong Q, Zhang J, Li Y. miR-26b is downregulated in human tongue squamous cell carcinoma and regulates cell proliferation and metastasis through a COX-2-dependent mechanism. Oncology reports. 2015;33:974-80.
[22] Verghese ET, Drury R, Green CA, Holliday DL, Lu X, Nash C, et al. MiR-26b is down-regulated in carcinoma-associated fibroblasts from ER-positive breast cancers leading to enhanced cell migration and invasion. The Journal of pathology. 2013;231:388-99.
[23] Saad Z, Arif M, Yaseen N, Jasim H, Jelawe M, Brown J. Novel expression of microRNAs in serum samples of Iraqi breast cancer women2014.
[24] Kumar S, Reddy PH. Are circulating microRNAs peripheral biomarkers for Alzheimer's disease? Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2016;1862:1617-27.
[25] Li H, Wu Q, Li T, Liu C, Xue L, Ding J, et al. The miR-17-92 cluster as a potential biomarker for the early diagnosis of gastric cancer: evidence and literature review. Oncotarget. 2017;8:45060-71.
[26] Satoh J-i, Kino Y, Niida S. MicroRNA-Seq Data Analysis Pipeline to Identify Blood Biomarkers for Alzheimer's Disease from Public Data. Biomarker Insights. 2015;10:BMI.S25132.
[27] Bai X, Han G, Liu Y, Jiang H, He Q. MiRNA-20a-5p promotes the growth of triple-negative breast cancer cells through targeting RUNX3. Biomedicine & Pharmacotherapy. 2018;103:1482-9.
[28] Zhao J, Yue D, Zhou Y, Jia L, Wang H, Guo M, et al. The Role of MicroRNAs in Aβ Deposition and Tau Phosphorylation in Alzheimer's Disease. Frontiers in neurology. 2017;8:342-.
[29] Li Y, Chen D, Jin L, Liu J, Su Z, Li Y, et al. MicroRNA-20b-5p functions as a tumor suppressor in renal cell carcinoma by regulating cellular proliferation, migration and apoptosis. Molecular medicine reports. 2016;13:1895-901.
[30] Ao X, Nie P, Wu B, Xu W, Zhang T, Wang S, et al. Decreased expression of microRNA-17 and microRNA-20b promotes breast cancer resistance to taxol therapy by upregulation of NCOA3. Cell Death &Amp; Disease. 2016;7:e2463.
[31] Wu Y, Xu J, Xu J, Cheng J, Jiao D, Zhou C, et al. Lower Serum Levels of miR-29c-3p and miR-19b-3p as Biomarkers for Alzheimer's Disease. The Tohoku journal of experimental medicine. 2017;242:129-36.
[32] Li N, Miao Y, Shan Y, Liu B, Li Y, Zhao L, et al. MiR-106b and miR-93 regulate cell progression by suppression of PTEN via PI3K/Akt pathway in breast cancer. Cell death & disease. 2017;8:e2796.
[33] Basavaraju M, de Lencastre A. Alzheimer's disease: presence and role of microRNAs. Biomolecular concepts. 2016;7:241-52.
[34] Zeng Q, Jin C, Chen W, Xia F, Wang Q, Fan F, et al. Downregulation of serum miR-17 and miR-106b levels in gastric cancer and benign gastric diseases. Chinese journal of cancer research = Chung-kuo yen cheng yen chiu. 2014;26:711-6.
[35] Denk J, Oberhauser F, Kornhuber J, Wiltfang J, Fassbender K, Schroeter ML, et al. Specific serum and CSF microRNA profiles distinguish sporadic behavioural variant of frontotemporal dementia compared with Alzheimer patients and cognitively healthy controls. PLOS ONE. 2018;13:e0197329.
[36] Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. Clinical cancer research : an official journal of the American Association for Cancer Research. 2013;19:4477-87.
[37] Dong H, Li J, Huang L, Chen X, Li D, Wang T, et al. Serum MicroRNA Profiles Serve as Novel Biomarkers for the Diagnosis of Alzheimer's Disease. Disease markers. 2015;2015:625659-.
[38] Pardo OE, Castellano L, Munro CE, Hu Y, Mauri F, Krell J, et al. miR-515-5p controls cancer cell migration through MARK4 regulation. EMBO reports. 2016;17:570-84.
[39] Pichler S, Gu W, Hartl D, Gasparoni G, Leidinger P, Keller A, et al. The miRNome of Alzheimer's disease: consistent downregulation of the miR-132/212 cluster. Neurobiology of Aging. 2017;50:167.e1-.e10.
[40] Orangi E, Motovali-Bashi M. Evaluation of miRNA-9 and miRNA-34a as potential biomarkers for diagnosis of breast cancer in Iranian women. Gene. 2019;687:272-9.
[41] Sarkar S, Jun S, Rellick S, Quintana DD, Cavendish JZ, Simpkins JW. Expression of microRNA-34a in Alzheimer's disease brain targets genes linked to synaptic plasticity, energy metabolism, and resting state network activity. Brain Res. 2016;1646:139-51.

[42] He Z, Yi J, Liu X, Chen J, Han S, Jin L, et al. MiR-143-3p functions as a tumor suppressor by regulating cell proliferation, invasion and epithelial–mesenchymal transition by targeting QKI-5 in esophageal squamous cell carcinoma. Molecular Cancer. 2016;15:51.
[43] Llorens F, Thune K, Andres-Benito P, Tahir W, Ansoleaga B, Hernandez-Ortega K, et al. MicroRNA Expression in the Locus Coeruleus, Entorhinal Cortex, and Hippocampus at Early and Middle Stages of Braak Neurofibrillary Tangle Pathology. Journal of molecular neuroscience : MN. 2017;63:206-15.

4.6.4 Table S2. Enrichment results of the miRNA family in the network using miRNet

Name	Hits	Pval
mir-515	15	2E-05
mir-17	6	0.0049
mir-6089	1	1
mir-4703	1	1
mir-4446	1	1
mir-3926	1	1
mir-3202	1	1
mir-3190	1	1
mir-1193	1	1
mir-3149	1	1
mir-1908	1	1
mir-670	1	1
mir-766	1	1
mir-1301	1	1
mir-629	1	1
mir-627	1	1
mir-623	1	1
mir-615	1	1
mir-548	3	1
mir-335	1	1
mir-124	1	1
mir-214	1	1
mir-204	2	1
mir-183	1	1
mir-181	1	1
mir-34	1	1
mir-7	1	1
mir-197	1	1
mir-103	2	1
mir-101	1	1
mir-10	1	1
mir-26	1	1
mir-24	1	1
mir-19	2	1
let-7	1	1

4.6.5 Table S3. Details of the miRNA network visualized in Figure6

The table shows the degree and betweenness value of the miRNA-gene network constructed using the query genes. The network was filtered on degree filter of 1 for miRNA nodes to reduce orphan branches of the miRNAs. This filtering limits to miRNAs connected with at least two nodes in the network.

Node	Degree	Betweeness
ZNF107	31	1883.72
NRAS	23	1365.81
ARAP2	20	969.054
PPP2CA	19	567.834
ZNF138	16	403.616
TOMM40	15	951.855
VDAC1	12	394.269
FSTL4	12	333.513
UBE2B	11	432.891
SKP1	7	135.951
hsa-mir-17-5p	4	161.301
hsa-mir-20a-5p	4	321.588
hsa-mir-335-5p	4	146.401
hsa-mir-19a-3p	3	64.9854
hsa-mir-19b-3p	3	64.9854
hsa-mir-93-5p	3	61.3012
hsa-mir-197-3p	3	60.1961
hsa-mir-34a-5p	3	221.191
hsa-mir-106b-5p	3	61.3012
hsa-mir-20b-5p	3	61.3012
hsa-mir-515-5p	3	119.672
hsa-mir-519e-5p	3	119.672
hsa-mir-519d-3p	3	61.3012
hsa-mir-183-3p	3	94.5203
hsa-mir-24-3p	2	69.4484
hsa-mir-26b-5p	2	100
hsa-mir-101-3p	2	45.092
hsa-mir-103a-3p	2	66.8742
hsa-mir-106a-5p	2	17.5695
hsa-mir-107	2	66.8742
hsa-mir-7-5p	2	32.0249

hsa-mir-181a-5p	2	133.059
hsa-mir-204-5p	2	13.69
hsa-mir-211-5p	2	13.69
hsa-mir-214-3p	2	41.4522
hsa-mir-124-3p	2	157.193
hsa-mir-125b-5p	2	29.3292
hsa-mir-519c-5p	2	9.93981
hsa-mir-518f-5p	2	9.93981
hsa-mir- 526a	2	9.93981
hsa-mir-524-5p	2	53.314
hsa-mir-520d-5p	2	53.314
hsa-mir-615-3p	2	78.7784
hsa-mir-548c-3p	2	17.2102
hsa-mir-623	2	13.69
hsa-mir-629-3p	2	6.63913
hsa-mir-766-3p	2	6.63913
hsa-mir-523-5p	2	9.93981
hsa-mir-518e-5p	2	9.93981
hsa-mir-522-5p	2	9.93981
hsa-mir-519a-5p	2	9.93981
hsa-mir-519b-5p	2	9.93981
hsa-mir-520c-5p	2	9.93981
hsa-mir-518d-5p	2	9.93981
hsa-mir-1301-3p	2	13.6568
hsa-mir-1908-5p	2	100
hsa-mir-2054	2	6.63913
hsa-mir-2681-5p	2	25.3689
hsa-mir-3123	2	25.3689
hsa-mir-548s	2	35.2714
hsa-mir-548u	2	9.93981
hsa-mir-3149	2	65.0465
hsa-mir-1193	2	103.964
hsa-mir-3202	2	24.9888
hsa-mir-3674	2	6.63913
hsa-mir-3685	2	186.615
hsa-mir-3692-3p	2	6.63913
hsa-mir-3926	2	35.2714
hsa-mir-4459	2	69.4484
hsa-mir-4468	2	133.059
hsa-mir-4533	2	24.9888
hsa-mir-3925-3p	2	6.63913
hsa-mir-4446-5p	2	13.69
hsa-mir-4640-5p	2	24.9888
hsa-mir-4703-5p	2	45.092

hsa-mir-4715-3p	2	28.9808
hsa-mir-4726-5p	2	24.9888
hsa-mir-4755-5p	2	13.69
hsa-mir-5006-3p	2	13.69
hsa-mir-5186	2	92.2074
hsa-mir-3190-3p	2	213.744
hsa-mir-98-3p	2	95.2022
hsa-mir-6087	2	68.9903
hsa-mir-6089	2	35.2714
hsa-mir-627-3p	2	29.9518
hsa-mir-670-3p	2	6.63913
hsa-mir-6832-5p	2	33.3635
hsa-mir-6845-3p	2	6.63913
hsa-mir-7106-5p	2	25.4254
hsa-mir-7111-3p	2	84.5759
hsa-mir-7161-5p	2	9.93981
hsa-mir-8087	2	89.387
MARK4	2	104.617
AREG	2	53.2203
BCAM	2	12.4426
HTRA3	2	81.2138
DTHD1	2	12.0013
CLPTM1	1	0
TCF7	1	0
RELB	1	0
APOE	1	0
APOC1	1	0

4.6.6 Table S4. miRNA target sites associated with query SNP- rs6859

Source: http://bioinfo.bjmu.edu.cn/mirsnp/search/

PVRL2	**hsa-miR-143-5p**	**6859**	**None**	**A==> T**	**create**	**A**	**None**	**None**	**None**	**T**	**167**	**-21.62**	**0.402**
PVRL2	hsa-miR-143-5p	6859	None	C==> T	create	C	None	None	None	T	167	-21.62	0.402
PVRL2	hsa-miR-143-5p	6859	None	G==> T	create	G	None	None	None	T	167	-21.62	0.402
PVRL2	hsa-miR-147b	6859	None	A==> C	create	A	None	None	None	C	150	-23.34	0.545
PVRL2	hsa-miR-147b	6859	None	C==> G	break	C	150	-23.34	0.545	G	None	None	None
PVRL2	hsa-miR-147b	6859	None	C==> T	break	C	150	-23.34	0.545	T	None	None	None
PVRL2	hsa-miR-199a-5p	6859	-0.321	A==> T	enhance	A	158	-20.18	0.993	T	166	-25.61	0.993
PVRL2	hsa-miR-199a-5p	6859	-0.321	C==> T	enhance	C	158	-20.23	0.993	T	166	-25.61	0.993
PVRL2	hsa-miR-199a-5p	6859	-0.321	G==> T	enhance	G	158	-21.3	0.993	T	166	-25.61	0.993
PVRL2	hsa-miR-199b-5p	6859	-0.323	A==> T	enhance	A	150	-17.77	0.993	T	158	-21.83	0.993
PVRL2	hsa-miR-199b-5p	6859	-0.323	C==> T	enhance	C	150	-17.77	0.993	T	158	-21.83	0.993
PVRL2	hsa-miR-199b-5p	6859	-0.323	G==> T	enhance	G	150	-17.93	0.993	T	158	-21.83	0.993
PVRL2	hsa-miR-584-3p	6859	None	A==> T	create	A	None	None	None	T	155	-25.83	0.008
PVRL2	hsa-miR-584-3p	6859	None	C==> T	create	C	None	None	None	T	155	-25.83	0.008
PVRL2	hsa-miR-584-3p	6859	None	G==> T	create	G	None	None	None	T	155	-25.83	0.008
PVRL2	hsa-miR-648	6859	-0.271	A==> T	enhance	A	143	-14.14	0.831	T	151	-19.94	0.831
PVRL2	hsa-miR-648	6859	-0.271	C==> T	enhance	C	143	-17.99	0.831	T	151	-19.94	0.831
PVRL2	hsa-miR-648	6859	-0.271	G==> T	enhance	G	143	-14.53	0.831	T	151	-19.94	0.831

4.6.7 Table S5. miR target site associated with query SNP- rs11556505

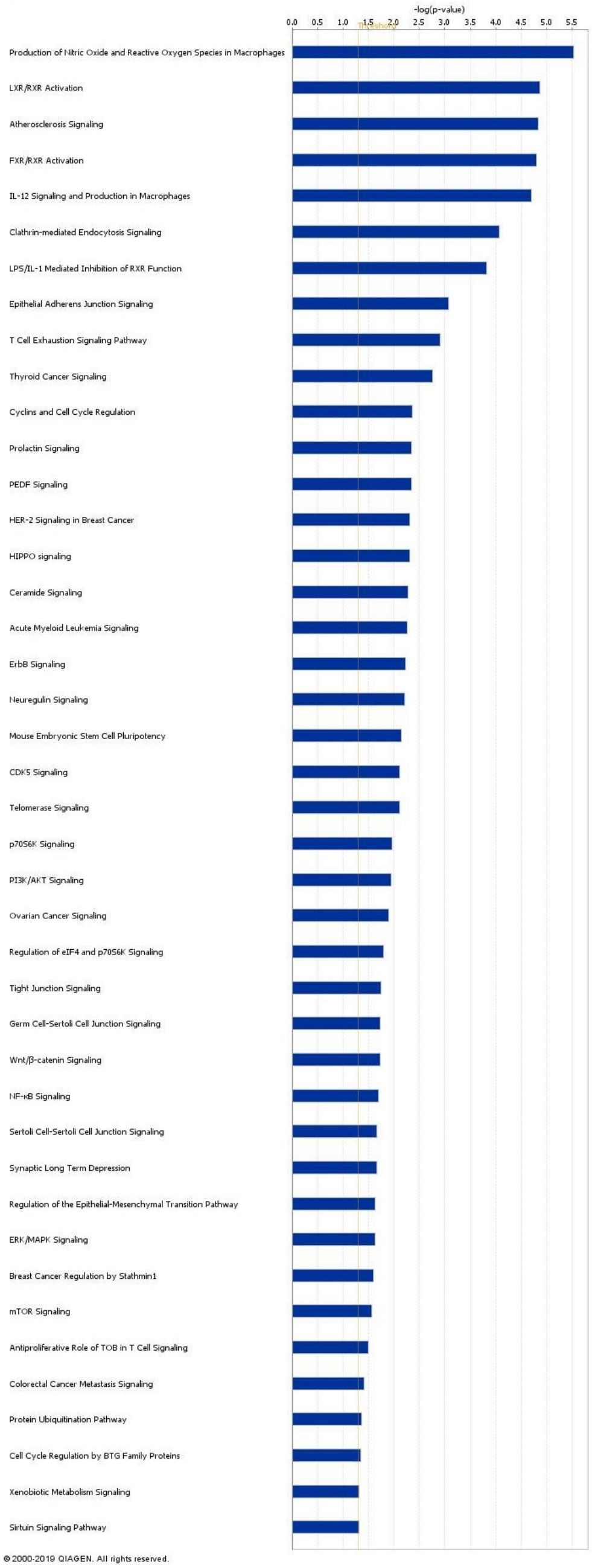
Source: http://compbio.uthsc.edu/miRSNP/

RefSeq ID:	**NM_001128916**
Gene Symbol:	TOMM40
Description:	Homo sapiens translocase of outer mitochondrial membrane 40 homolog (yeast) (TOMM40), nuclear gene encoding mitochondrial protein, transcript variant 2, mRNA.

Gene Location:	chr19(+):45394476-45406946

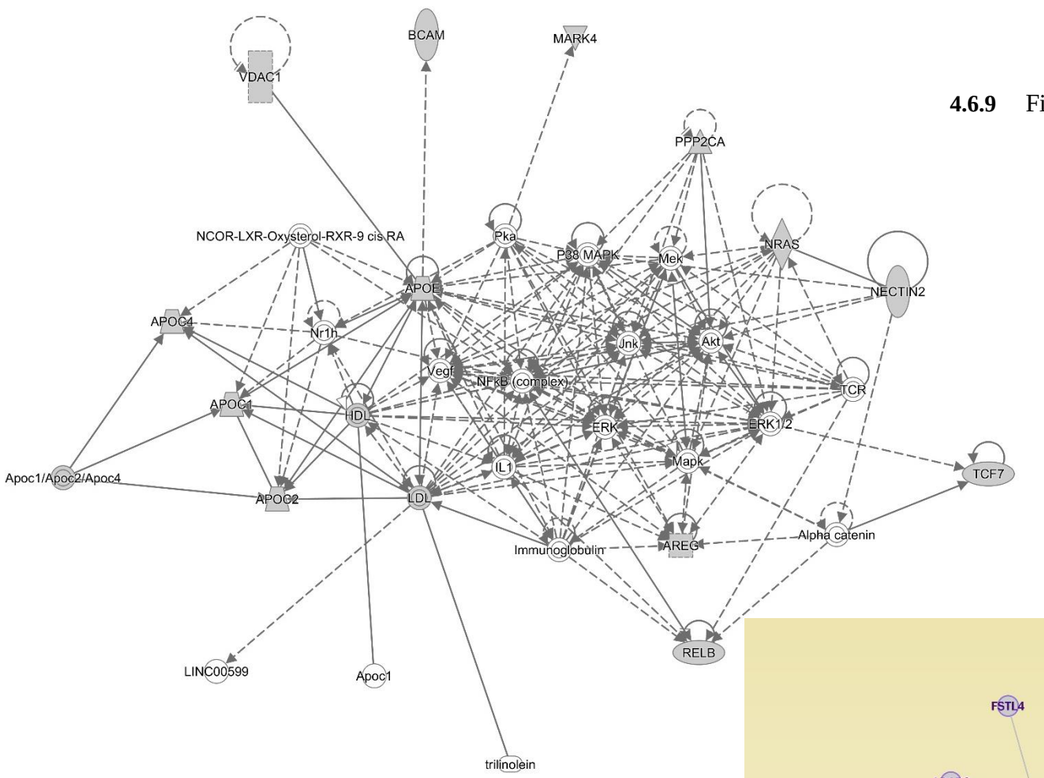
SNPs and INDELs in miRNA target sites from CLASH data				
CLASHSequenceID	**miRNA**	**dbSNP ID**	**miRSite**	**Strand**
L2HS-340749_4	hsa-miR-484	rs11556505	actT**C**GGGGtcacatatgtGGGGACaaAGCagCTGAgtcccacagaggcgttccc	+

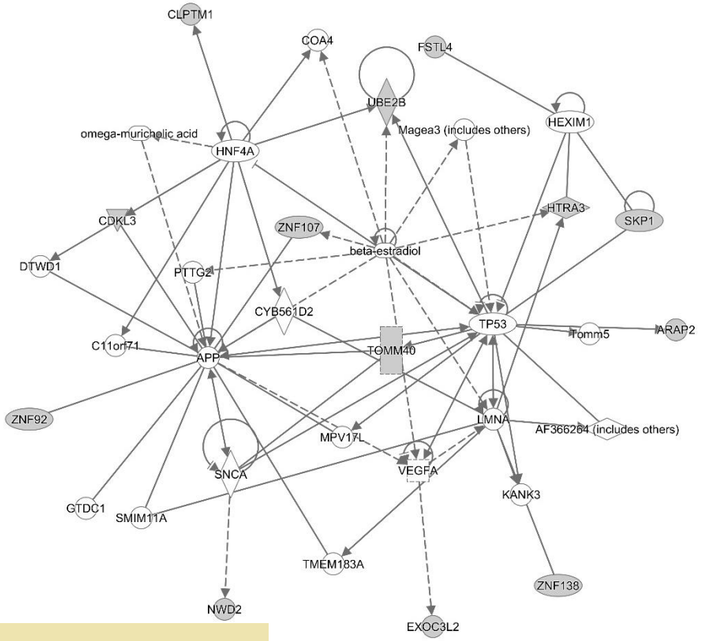
4.6.8 Figure S3. Enriched processes in reported genes



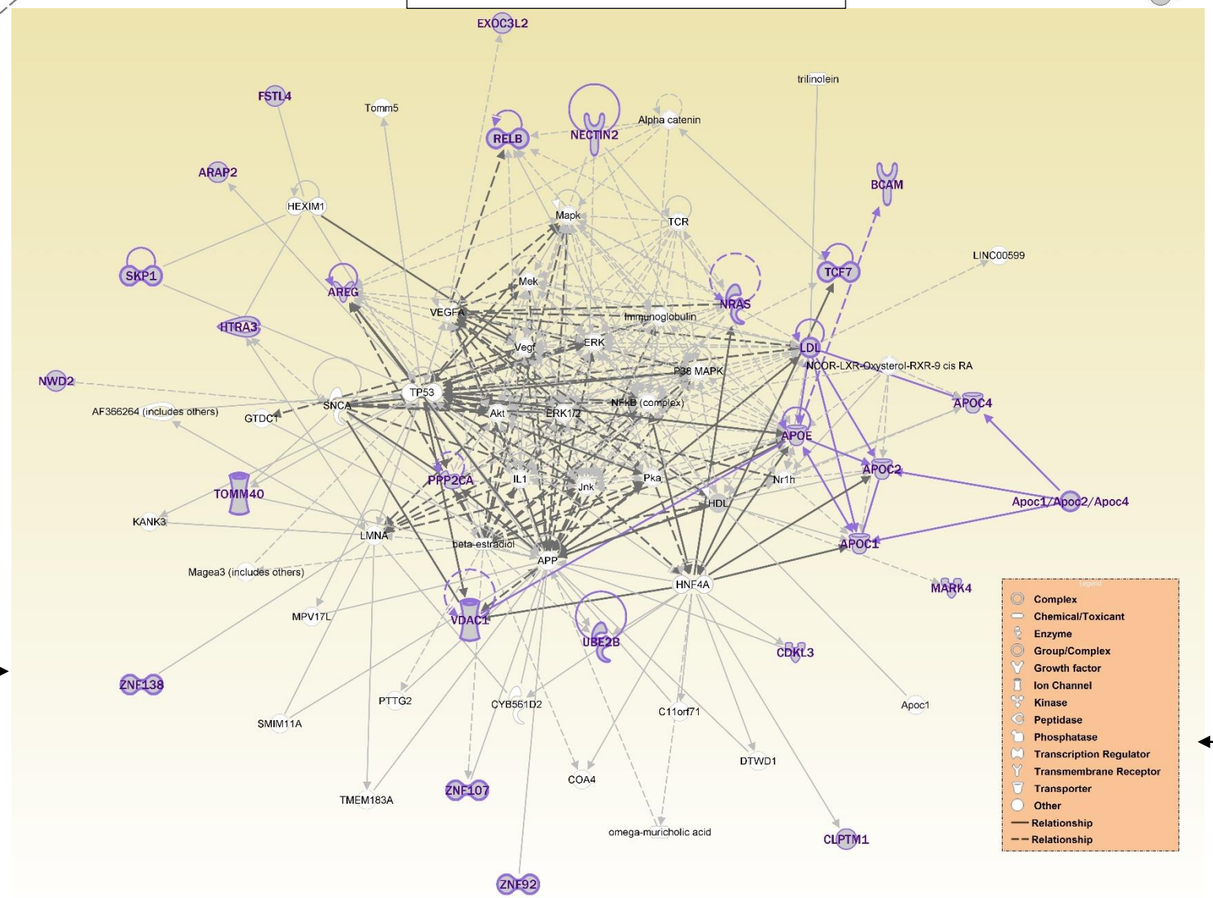
4.6.9 Figure S4. Network analysis of reported genes

Network analysis in IPA software using the classified significant genes in two networks. We merged these two networks to analyze chief processes involved. For each network, few of the top canonical pathways are displayed under the network.





Merged Network



Network 1
Lipid Metabolism

Pathway Name	# Molecules	Molecule(s)
ILK Signaling	10	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NFkB...
Renin-Angiotensin Signaling	10	ERK, ERK1/2, HDL, Jnk, Mapk, Mek, NFk...
IL-15 Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NFkB...
PKC8 Signaling in T Lymphocytes	9	ERK, ERK1/2, Jnk, Mapk, Mek, NFkB (co...
ErbB Signaling	9	Akt, AREG, ERK, ERK1/2, Jnk, Mapk, Me...
Role of Pattern Recognition Receptors in...	9	ERK, ERK1/2, HDL, IL1, Jnk, Mapk, Mek,...
Mouse Embryonic Stem Cell Pluripotency	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NRA...
Role of NFAT in Cardiac Hypertrophy	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NRA...
IL-1 Signaling	9	ERK, ERK1/2, IL1, Jnk, Mapk, Mek, NFkB...
Regulation of IL-2 Expression in Activate...	9	ERK, ERK1/2, Jnk, Mapk, Mek, NFkB (co...
Role of MAPK Signaling in the Pathogene...	9	Akt, ERK, ERK1/2, HDL, Jnk, Mapk, Mek,...
Type II Diabetes Mellitus Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NFkB...
Fc Epsilon RI Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NRA...
G-Protein Coupled Receptor Signaling	9	Akt, ERK, ERK1/2, Mapk, Mek, NFkB (co...
T Cell Receptor Signaling	9	ERK, ERK1/2, Jnk, Mapk, Mek, NFkB (co...
Axonal Guidance Signaling	9	Akt, ERK, ERK1/2, Mapk, Mek, NRAS, P3...
Ceramide Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NFkB...
PI3K/AKT Signaling	9	Akt, ERK, ERK1/2, Mapk, Mek, NFkB (co...
Nitric Oxide Signaling in the Cardiovascul...	9	Akt, ERK, ERK1/2, HDL, Mapk, Mek, P38 ...
Type I Diabetes Mellitus Signaling	9	ERK, ERK1/2, IL1, Jnk, Mapk, Mek, NFkB...
Systemic Lupus Erythematosus Signaling	9	Akt, ERK, ERK1/2, HDL, IL1, Mapk, NRA...
Neuroinflammation Signaling Pathway	9	Akt, ERK, ERK1/2, IL1, Jnk, Mapk, Mek, ...
STAT3 Pathway	9	ERK, ERK1/2, IL1, Jnk, Mapk, Mek, NRA...
GNRH Signaling	9	ERK, ERK1/2, Jnk, Mapk, Mek, NFkB (co...
Endocannabinoid Developing Neuron Pat...	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NRA...
Pancreatic Adenocarcinoma Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NFkB...
NGF Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NFkB...
AMPK Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, P38 ...
IL-17 Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NFkB...
Role of NFAT in Regulation of the Immun...	9	Akt, ERK, ERK1/2, Mapk, Mek, NFkB (co...
IGF-1 Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NRA...
Insulin Receptor Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NRA...
CDK5 Signaling	9	ERK, ERK1/2, Jnk, Mapk, Mek, NRAS, P3...
mTOR Signaling	9	Akt, ERK, ERK1/2, Mapk, NRAS, P38 MA...
Thrombin Signaling	9	Akt, ERK, ERK1/2, Jnk, Mapk, Mek, NFkB...

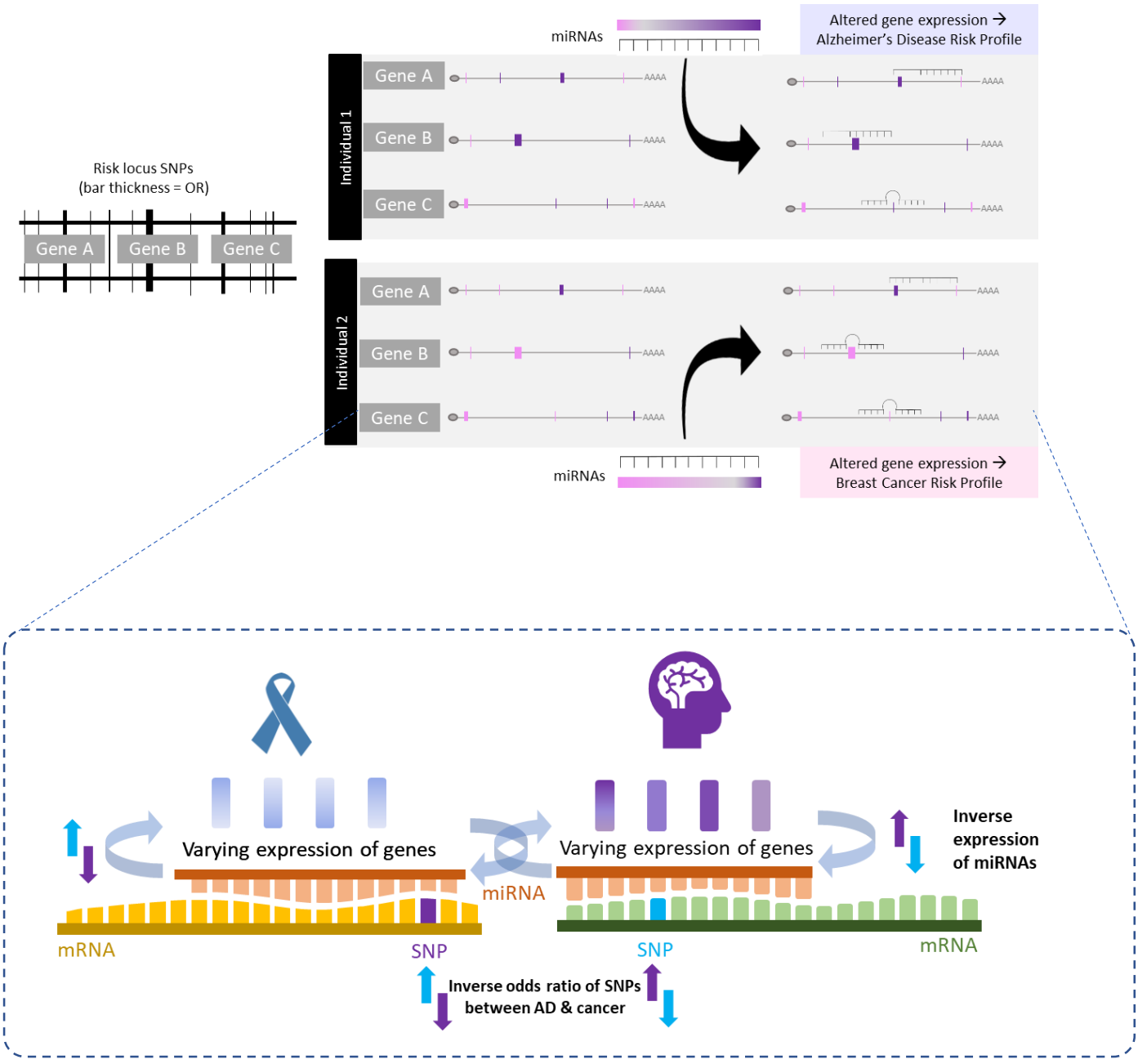
Network 2
Cell Death and Survival

Pathway Name	# Molecules	Molecule(s)
Sirtuin Signaling Pathway	4	APP, Tomm5, TOMM40, TP53
Hypoxia Signaling in the Cardiovascular ...	3	TP53, UBE2B, VEGFA
Neuroinflammation Signaling Pathway	2	APP, SNCA
Pancreatic Adenocarcinoma Signaling	2	TP53, VEGFA
Cell Cycle: G2/M DNA Damage Checkpoin...	2	SKP1, TP53
Protein Ubiquitination Pathway	2	SKP1, UBE2B
Neuroprotective Role of THOP1 in Alzhei...	2	APP, HTRA3
Colorectal Cancer Metastasis Signaling	2	TP53, VEGFA
Role of Tissue Factor in Cancer	2	TP53, VEGFA
Bladder Cancer Signaling	2	TP53, VEGFA
Cyclins and Cell Cycle Regulation	2	SKP1, TP53
Ovarian Cancer Signaling	2	TP53, VEGFA
Apoptosis Signaling	2	LMNA, TP53
Sumoylation Pathway	2	SNCA, TP53
Inhibition of Angiogenesis by TSP1	2	TP53, VEGFA
Amyotrophic Lateral Sclerosis Signaling	2	TP53, VEGFA
Cell Cycle: G1/S Checkpoint Regulation	2	SKP1, TP53
Huntington's Disease Signaling	2	SNCA, TP53
HIF1α Signaling	2	TP53, VEGFA
Mitochondrial Dysfunction	2	APP, SNCA
Maturity Onset Diabetes of Young (MOD...	1	HNF4A
UVC-Induced MAPK Signaling	1	TP53
UVB-Induced MAPK Signaling	1	TP53
Role of CHK Proteins in Cell Cycle Check...	1	TP53
UVA-Induced MAPK Signaling	1	TP53
Role of BRCA1 in DNA Damage Response	1	TP53
STAT3 Pathway	1	VEGFA
Myc Mediated Apoptosis Signaling	1	TP53
Endocannabinoid Cancer Inhibition Path...	1	VEGFA
Role of PKR in Interferon Induction and ...	1	TP53
IL-8 Signaling	1	VEGFA
Mouse Embryonic Stem Cell Pluripotency	1	TP53
NGF Signaling	1	TP53
14-3-3-mediated Signaling	1	SNCA
eNOS Signaling	1	VEGFA
Transcriptional Regulatory Network in E...	1	HNF4A

4.6.10 Figure S5. Potential mechanism of SNPs-miRNA binding sites

Simplified schematic of the plausible relation between SNPs with cross-phenotypic effects and miRNAs exerting inverse alterations in the two phenotypes. The panel on the left represents the variability of odds ratio of SNPs within multiple genes. The panel on the right represents the different combinations of varying effect sizes and overlapping genes within two individuals thus altering their miRNA expression levels resulting in inverse phenotypes.

The relationship between miRNAs and gene expression is complex. Changes in nucleotide variation lead to untargeted binding and miRNAs with similar sequences compensate for misguided binding by relying on codon degeneracy. The variation of different gene expression possibly affects the same miRNA in a cyclical mode. The reported miRNAs require functional studies characterizing their expression, targets and phenotypic consequences.



Chapter 5

SNP-DERIVED TRANSCRIPTOMICS OF MULTIPLE TISSUES BETWEEN ALZHEIMER'S AND CANCER IMPLICATES MIR-17 FAMILY AND SIRTUIN SIGNALING

ABSTRACT. Alzheimer's disease (AD) and cancer are two aging-associated diseases, and several epidemiological findings have reported an inverse correlation between the two diseases. Even though aging is a risk factor for Alzheimer's and cancer, the two diseases have been known to have discrete molecular mechanisms. The goal of this study is to identify genetic variants that may be influencing functional changes on Alzheimer's disease and cancer biology. We used individual-level genotype data to predict transcriptome of five brain tissues for AD-specific genes and breast/prostate tissue for respective cancers using PrediXcan. A total of 6279 males were used by comparing common control males between AD and prostate cancer and 3359 females between AD and breast cancer against common control females. All the subjects were limited within 60-80 years of age to remove heterogeneity from age. Several of the genes associated with both AD and cancer are inversely expressed based on the z-scores of the association.

These overlapping genes were enriched for sirtuin signaling and miRNA-17 and miRNA-15 family. These findings were also reported in our previous study involving variants with inverse odds ratios. Based on our results, we hypothesize that p53, mitochondrial dysregulation and miRNAs contribute to differences in type-2 and type 3 EMT in AD and cancer.

Short Title: SNP-derived transcriptomics between Alzheimer's & cancer

5.1 INTRODUCTION

With the rise in the aging population, prevalence of two diseases – Alzheimer's and cancer continue to increase mortality and morbidity rates in the US[1]. Alzheimer's disease is a neurodegenerative disorder which affects 50 million people worldwide the number is expected to increase up to 130 Million by 2030 due to increase in life expectancy in the aging population (65 years or older)[2]. The AD symptomology includes behavioral difficulties, progressive cognitive and memory decline and speech impediments[2]. AD pathology includes neuronal loss, dysfunction of the synapse and plaque formation composed of a beta amyloid and/or neurofibrillary tangles. In the aging population sporadic AD is more common with multifactorial etiology which has been difficult to ascertain so far[2]. On the other hand, two most prevalent cancers in the aging population are breast and prostate cancer. It is estimated that approximately 6 million individuals will have a history or breast or prostate cancer, due to the increase in survival rates from better treatment[3]. Cancer refers to a group of diseases wherein a body cell loses apoptotic regulation leading to malignant traits and spreading into or invading nearby tissues[4]. The cellular deterioration from aging is known to cross paths with both cancer and Alzheimer's disease causing a time-dependent accumulation of cellular damage[5]. In the last decade, epidemiological findings have shown that Alzheimer's and cancer share a negative correlation or inverse comorbidity defined as lower than expected likelihood of developing a secondary disease after being diagnosed with the primary condition[6]. Longitudinal studies involving elderly people ages 65 and above have reported that cancer was less prevalent in individuals with AD (5.8%) compared to those who did not have any form of dementia (26.5%)[7]. These findings were later supported by a retrospective study of more than three and a half million US veterans age 65 and above which confirmed a lower risk of Alzheimer's disease in cancer survivors[8].

These epidemiological observations have fueled numerous systematic reviews and hypothesis-driven investigations to understand the common and different characteristics of both diseases. Transcriptomic meta-analysis studies have shown the inverse regulation of *Pin1*, *p53* and Wnt signaling, dysregulation in mitochondrial metabolism and oxidative phosphorylation in AD and cancer [9]. Additionally, factors such as *APOE4* which are high risk factors for late-onset AD have been shown to be protective in cancer in the Framingham Heart Study. The role of microRNA's is crucial for the regulation of gene expression, and miRNAs such as miR-34, miR-9 and miR-17 have been found to be inversely expressed in AD and cancer[10]. Our previous study involving individual level SNP effect between AD and cancer confirmed several of the hypothesized and reported findings from other studies. In this study we want to understand the relationship of transcriptomics between AD and cancer using SNP derived gene expression. SNPs in regulatory or coding region are highly correlated with tissue – specific alterations in gene expression[11]. These correlation weights have been estimated in the PrediXcan[11] model using reference datasets such as the GTEx project which contains RNA-seq and genotype information of 53 tissues from 620 donors (v7)[12]. Approximately, 80% of the individuals in the GTEx dataset are of European ethnicity and more than 65% of the individuals are over the age of 50 years[12]. There is limited overlap in genetic factors that contribute to early and late-onset forms of AD and cancer[13]. To reduce genetic variance from age, we compared individuals between 60-80 years of age with late – onset AD or cancer. The use of SNP-derived transcriptomics offers the advantage of comparing multiple brain tissues in the same set of control population, as opposed to meta-analysis of gene expression repositories, which may induce study-dependent bias, and limit harmonization of same tissue type. In this study, we compared tissue-specific gene expression profile of five brain tissues – cerebellum, cerebellar hemisphere, cortex, hypothalamus and hippocampus in controls vs AD,

against cancer tissues – breast or prostate tissue in controls vs cancer. To study the relationship of genes expression between AD and cancer, we compared genes that overlap between genes that were associated with AD - control vs AD (brain tissues) and prostate cancer - control vs cancer (prostate tissue). We then, compared the z-scores to identify which genes were upregulated or downregulated in each disease.

5.2 METHODS & MATERIALS

5.2.1 Data description – ADGC, ADNI, BPC3

The individual-level genotype and clinical information was obtained from ADGC (phs000372.v1.p1), and Breast Prostate Cancer Care Consortium (BPC3) (phs000812.v1.p1) and Alzheimer's Disease Neuroimaging Initiative (ADNI) (www.adni-info.org). These datasets were chosen because they all were genotyped on Illumina Human660W-Quad, to minimize any technical bias and harmonization issues while merging the datasets. The research protocol for this project was reviewed by University of North Texas Health Science Center Institutional Review Board on June 24, 2016 and determined to be exempt human subject research under IRB–2016-090.

5.2.2 Quality Control

We merged individuals from all datasets and performed two-level quality control procedure using plink v1.9[14] as per the guidelines by Anderson et al[15]. Individuals were removed based on failing heterozygosity or missing markers, IBD, and outliers on principal components PC1 and PC2. The individuals were of self-described European ethnicity, which was confirmed by comparing with the 1000 genome dataset. Controls who had family history of cancer, and all individuals with medical history of cancer in AD population were also removed. Followed by including individuals

within 60-80 years of age. SNP markers were removed based on missingness, differential call rate between cases and controls (1-e8), MAF (0.01) and HWE (1e-5). The final QC'd dataset was then stratified by sex (6279 males and 3359 females). This allows to compare same control population obtained from combining both AD and cancer cohorts against each phenotype.

5.2.3 Gene Expression Association

We used PrediXcan[11] to impute gene expression using individual level SNP profiles for males and females. Tissue specific SNP-gene expression weights trained using lasso regression from GTEx dataset (v7)[12], available at http://predictdb.org/. We downloaded model files for brain tissues - cerebellum, cerebellar hemisphere, cortex, hypothalamus and hippocampus, for identifying AD related changes and tissues – breast and prostate for studying cancer specific changes. We choose the following brain tissues because have been implicated in the etiology of AD and each tissue has more than 100 donors with genotype and RNA-seq data in the GTEx dataset (Tissue - No. of individuals; cerebellum – 154, cerebral hemisphere - 125, cortex-136, hippocampus -111 and hypothalamus 108). The breast tissue has 251 donors and the prostate tissue has 132 donors when genotype data. We predicted 4330 genes for cortex, 4758 genes for cerebral hemisphere, 6094 genes for cerebellum, 2817 genes for hippocampus and 2836 genes for hypothalamus, the association tests were performed for comparison of control vs AD (separately for males and females) in all brain tissues to identify AD-specific pathology. In females, we predicted 5308 genes for breast tissue and 3263 genes for prostate tissue in males. The z-scores identify the direction of up or down regulation the significant genes (p- value < 0.05). For the next analysis, we performed a cross tissue comparison to understand how genes present in both tissues – brain and cancer-related, may be playing different roles in pathology (AD/cancer) specific tissue-based expression. We identified overlapping significant genes between each brain tissue (control vs AD) and

breast/prostate tissue (control vs cancer). To remove any tissue-based bias we added a threshold that control vs cancer population should not be significant for brain tissues and control vs AD population should not be significant for cancer tissue (breast /prostate).

5.2.4 Enriched processes and pathways

To identify processes that maybe in opposite directions in AD and cancer, significant genes that were overlapping in the cross-tissue comparison between AD and cancer were analyzed in Ingenuity Pathway Analysis® (IPA) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) for canonical pathways and GeneMania[16].

5.2.5 miRNA family enrichment

The overlapping genes between AD and cancer were investigated further to identify associated miRNAs using miRNet[17]. The network was filtered on betweenness centrality of 1.0 for miRNA nodes, to prioritize miRNAs that have at least two or more gene targets in the network. All the miRNAs in the network were tested for enrichment of miRNA family using hypergeometric test (FDR; p-value <0.05).

5.3 RESULTS

5.3.1 AD & prostate cancer in males

To understand which genes are associated with both AD and prostate cancer in males, we compared genes that were overlapping in each of the five brain tissues (control vs AD) with prostate tissue (control vs cancer). Comparison between cerebellar hemisphere and prostate tissue, identified 15 genes. Most of these genes are inversely regulated (as per the z-score direction) between the two the diseases, except *DPYSL4*, *KRT18P34*, *PNRC2*, RP11-314C16.1, *SPEG* and

WTAP. We found 16 genes intersecting cerebellum and prostate tissue, out of which only four genes – *DPYSL4, MLLT10P1, NTPCR* and *VSIG8* were found to be in the same direction. Comparing genes between AD and cancer in cortex and prostate tissue resulted in 16 genes coinciding between the two states. Here, we see similar pattern, of less than 25% genes in the same direction in both AD and cancer – *HLA-DQA2, LRPAP1, PIGN* and RP11-757G1.6 (lnc-RNA). Comparison between hippocampus and hypothalamus with prostate tissue, identified 15 and 7 genes respectively (Figure 1 and 2).

These genes were further investigated for co-expression, pathway and genetic interactions. Several of the genes were found to be co-expressed, and genes such as the *TUBA1C*, *LRPAP1* were found to share pathways with lipid metabolism genes and inflammatory domains of HLA-gene family (Figure 3). All the genes were also tested for enriched pathways and signaling networks. The most representative pathway was found to be sirtuin signaling and comparing the gene expression direction for both AD and prostate cancer, we observe that most of the genes in the network are inversely expressed (Figure 4).

We also tested genes for overrepresented miRNAs, by constructing network that prioritized miRNAs that mapped to two or more genes. The network was enriched for miR-17 family (6 members). We also miRNAs such as miR-26b, miR-335, and 3 miRNAs from let-7 family (Figure 5). We found the enrichment of same miRNA family in our previous study involving identification of cross-phenotypic SNPs between AD and cancer targeting different set of genes than identified here.

5.3.2 AD & breast cancer in females

Upon comparing genes between brain and breast tissues for AD and breast cancer in females, we found four significant hits for each pairwise tissue comparison. We found four hits in cerebellar

hemisphere vs breast tissue – *CEP63, DTD1, ENTPD7* and *NPIPL1*. Three genes and one lncRNA was found to intersect cortex and breast tissue – *C6orf201, CTC-523E23.11, HNF1A* and RP11-757G1.6. Comparing hippocampus and breast tissue identified *CERS5, EPCAM, PFKL* and *ZNF147*. None of the genes for cerebellum and hypothalamus were found to be significant in the tissue-based comparisons (Figure 6).

Due to the limited number of genes between AD and breast cancer, we couldn't identify enriched processes. The genes were further as targets for miRNAs, the network was enriched for miRNA-17 (5 members) and miR-15 family (3 members). The members of miR-17 family target two genes – *ENTPD7* and *ZNF417*, both of which are inversely regulated in AD and cancer. The members of miR-15 family target *CEP63, DTD1, HNF1A* and *ENTPD7* (Figure 7).

5.4 DISCUSSION

AD and cancer have different etiologies. The "hallmarks of cancer" are highly dependent on metabolic alterations resulting in type 3 epithelial-to-mesenchymal transition (EMT)[18], in contrast, the AD pathology is a consequence of type-2 EMT caused due to inflammation from metabolic dysregulation in the brain [19]. The sirtuin homologues are members of the histone deacetylase family, affecting chromatin regulation and transcription factors[20]. Upon comparing brain transcriptomic profile of AD individuals and prostate transcriptome of cancer individuals with the same set of controls, we observe that most overlapping genes between the two diseases are inversely expressed. These genes are involved in sirtuin signaling, which is known to play role in both AD[2] and cancer[21].

Sirtuins group of proteins (*SIRT*1-7) are histone and chromatin transformers involved in regulation of oxidative stress, cellular metabolism and differentiation. SIRT-1,2, 6 and 7 are localized in nucleus and the nucleolus, and SIRT-3-5 are present in the mitochondria[20]. While not all sirtuins have been investigated in AD pathology, *SIRT-1-3* and 6 have been associated with AD. SIRT-1 is under expressed in AD brains and increasing the SIRT-1's levels is reported to be neuroprotective, as it is inversely correlated with Aβ and tau accumulation[22]. The overexpression of SIRT-1 is associated with downregulation of p53, increased AKT and CREB deactivation alleviating cognitive response in animal models[23]. Forkhead transcription factor regulates brain metabolism via interaction of Wnt-signaling and mitochondrial biogenesis by interacting with SIRT-1[20]. In cancer, SIRT-1 is found to be both under and overexpressed. The overexpression of SIRT-1 is found to deacetylate both of its tumor suppressing targets – p53 and E2F1 thus contributing to oncogenesis. The overexpression has also been known to regulate β-catenin signaling which induces DNA repair and protects from age-associated cancer development. The over expression of SIRT-1 has been reported in prostate cancer and associated with cancer progression[21] . SIRT2 is increased in patients with AD brains and known to cause tau aggregation by regulating mitochondrial and redox homeostasis [24].

One of the SNPs in *SIRT-2* (rs10410544_C) has been documented to increase predisposition to AD in APOE4 negative, and the T allele is more prevalent in APOE4 positive individuals with AD. *SIRT-2* interacts with both *APOE* and *CYPD26* resulting in pathogenic and therapeutic implications. Individuals with rs10410544-C/T and extensive metabolizers based on *CYPD26* have been identified to respond better to AD therapeutics over other carriers[20].

SIRT-2 is downregulated in multiple cancer types, including breast and prostate cancer. This under expression leads to prevention of mitotic cell cycle and reduced activity of anaphase-promoting

complex resulting in uncontrolled cellular proliferation. *SIRT-2* is elevated under oxidative stress which rescues mitotic dysregulation by inducing cell cycle checkpoint [21].

Mitochondrial sirtuins - *SIRT-3-5* regulate metabolic and mitochondrial energy based on availability of ATP. *SIRT-3* is reported to be either deleted or under expressed in cancer, however in hereditary ovarian cancer, *SIRT-3* is duplicated and suspected to play tumorigenic role. SIRT-3 dysregulation has been reported in cerebral cortex of AD brains in humans leading to neuronal and mitochondrial decline mediated by p53. *SIRT-4* is also reduced in multiple cancer, unlike *SIRT-5* which is overexpressed in lung cancer[21].

The under expression of SIRT-6 in AD patients is negatively correlated with DNA damage induced by accumulation of Aβ[20]. In cancer, SIRT-6 is found to be deficient in 20% of cancers and upregulated in breast cancer acting intervening stronger DNA damage response causing resistance to cancer treatment[21].

The micro-RNAs are related molecules consisting of 22nt of non-coding RNA sequence that bind to 3'UTR of the gene, controlling the transcriptional regulation of the gene[25]. We found overrepresentation of miR-17 family in both cancer to AD comparison, and representation of miR-15 family for genes overlapping between AD and breast cancer. The miR-17 family consists of miR-20a/b, 106a/b and 93[25], and the miR-15 family consists of miR-15a/b and miR-16[26]. Although the seed sequences of miR-16-1 and 16-2 are identical, miR-15a and 15b have different seed sequences and may contribute towards different functions[26]. The miR-15a/16-1 are tumor-suppressors and are deficient in multiple cancers[27]. This deficiency in their expression upregulates anti-apoptotic proteins such as Mcl-1 and promote tumorigenic potential by inducing EMT. The miR-15b/16-2 interacts with cyclin proteins, insulin growth like receptors and is also known to play anti-oncogenic role[26]. The miR-15b is over expressed in AD pathology, and lowering its

expression alleviates Aβ accumulation and reduces apoptotic activity[28]. In contrast, the miRNA-17 is found to be downregulated in AD and is inversely correlated with APP and Aβ expression[29]. The miRNA-17 is over expressed in cancers and is accompanied by surge in cellular proliferation of cancer tissues[30].

In our previous study, we reported the involvement of *TOMM40* and *VDAC1* in the inverse relationship of AD and cancer. These mitochondrial proteins are known to interact with sirtuin proteins, primarily the *SIRT-1*[31]. We also found the enrichment of miRNA-17 family and reported its opposite biological roles AD and cancer. In this study, we extended the previous investigation to confirm role of mitochondrial and metabolic dysfunction along with role of miRNAs in the transcriptomic profiles of AD and cancer. This study helped identify sirtuins as potential regulatory networks which play in dual roles in AD and cancer. Furthermore, we found statistical enrichment of miRNA-17 family in both cancers against AD. It is interesting to note that members of sirtuin homologue play dual roles, similar to members of miRNA-17. These findings highlight the role of biological degeneracy built into our aging biology. Investigation using animal models are warranted to tease apart the functional consequences of each homologue and evaluate the role of compensation provided by structurally similar molecules.

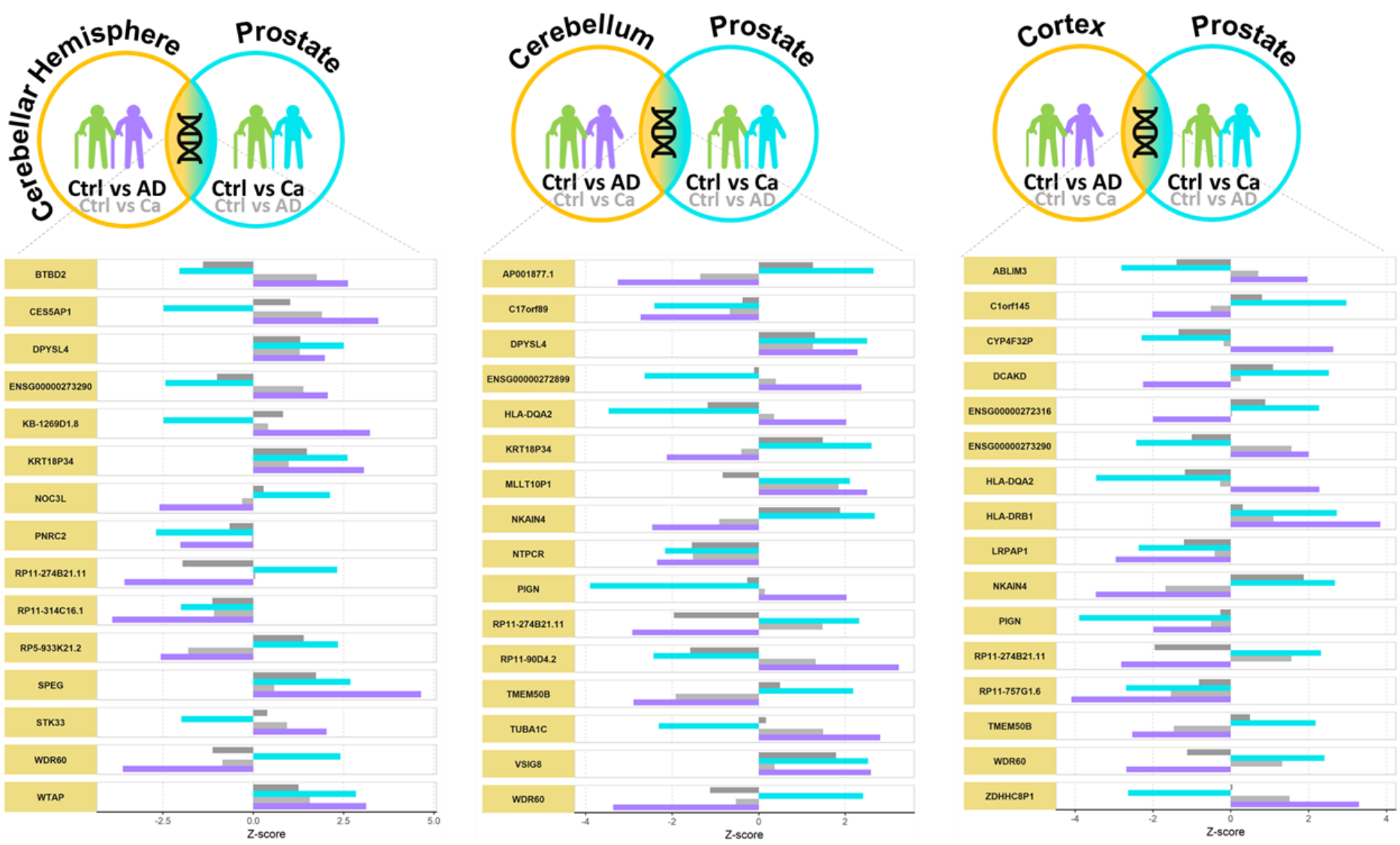


Figure 1: ***Transcriptomic profile of AD and prostate cancer****. The genes that were associated with AD in brain tissues and with prostate cancer in prostate tissue are listed here, derived from each pairwise comparison of brain and cancer tissue. The z-score identify if the gene is up/downregulated. Blue bars show significant association for ctrl vs cancer in males and purple bars show significant association for ctrl vs AD males. Grey bars represent that these genes were not significant between ctrl vs cancers in brain tissue and ctrl vs AD were not significant in prostate tissue. Ctrl – control; AD – Alzheimer's Disease; Ca - cancer*

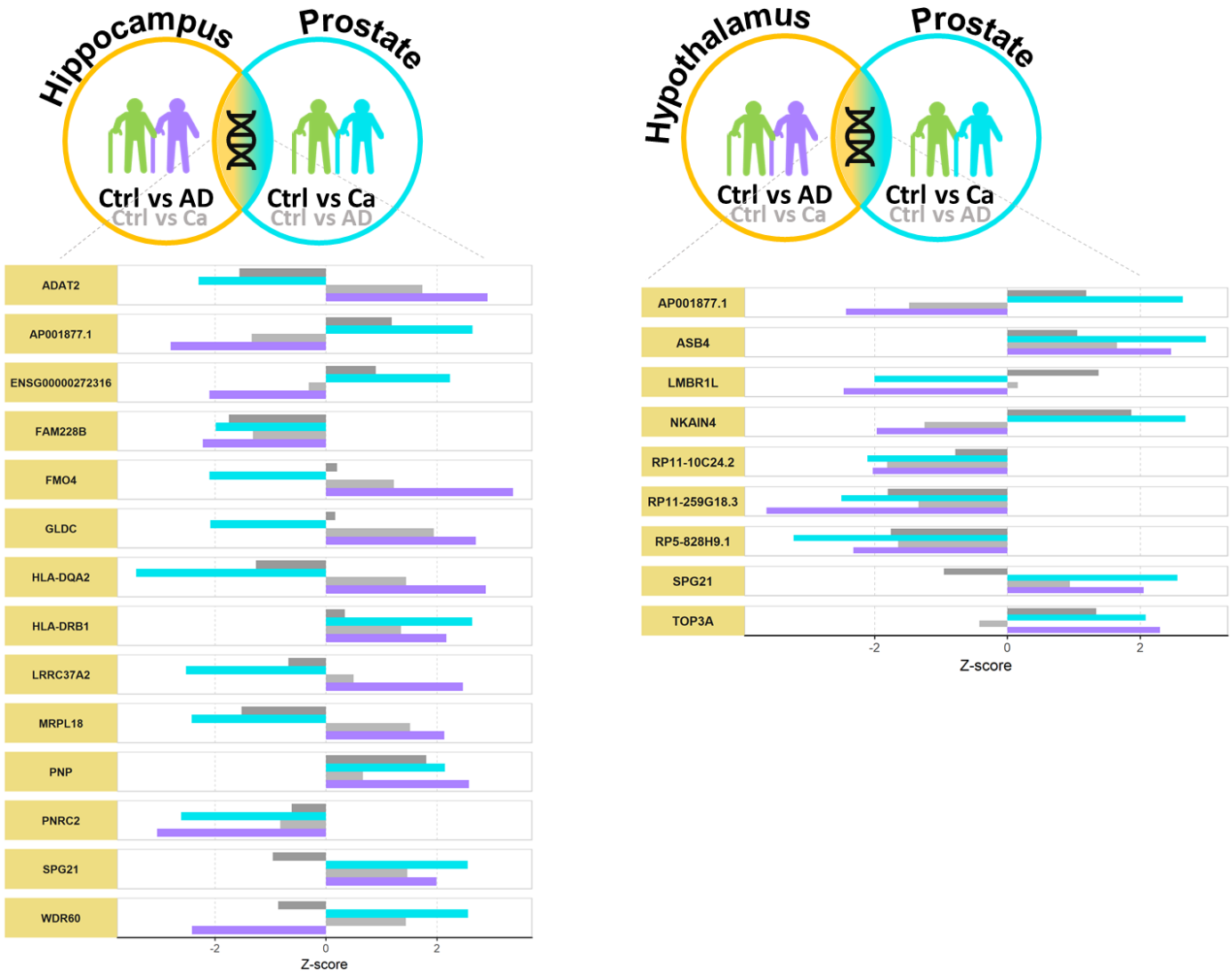


Figure 2: ***Transcriptomic profile of AD and prostate cancer****. The genes that were associated with AD in brain tissues and with prostate cancer in prostate tissue are listed here, derived from each pairwise comparison of brain and cancer tissue. The z-score identify if the gene is up/downregulated. Blue bars show significant association for ctrl vs cancer in males and purple bars show significant association for ctrl vs AD males. Grey bars represent that these genes were not significant between ctrl vs cancers in brain tissue and ctrl vs AD were not significant in prostate tissue. Ctrl – control; AD – Alzheimer's Disease; Ca - cancer*

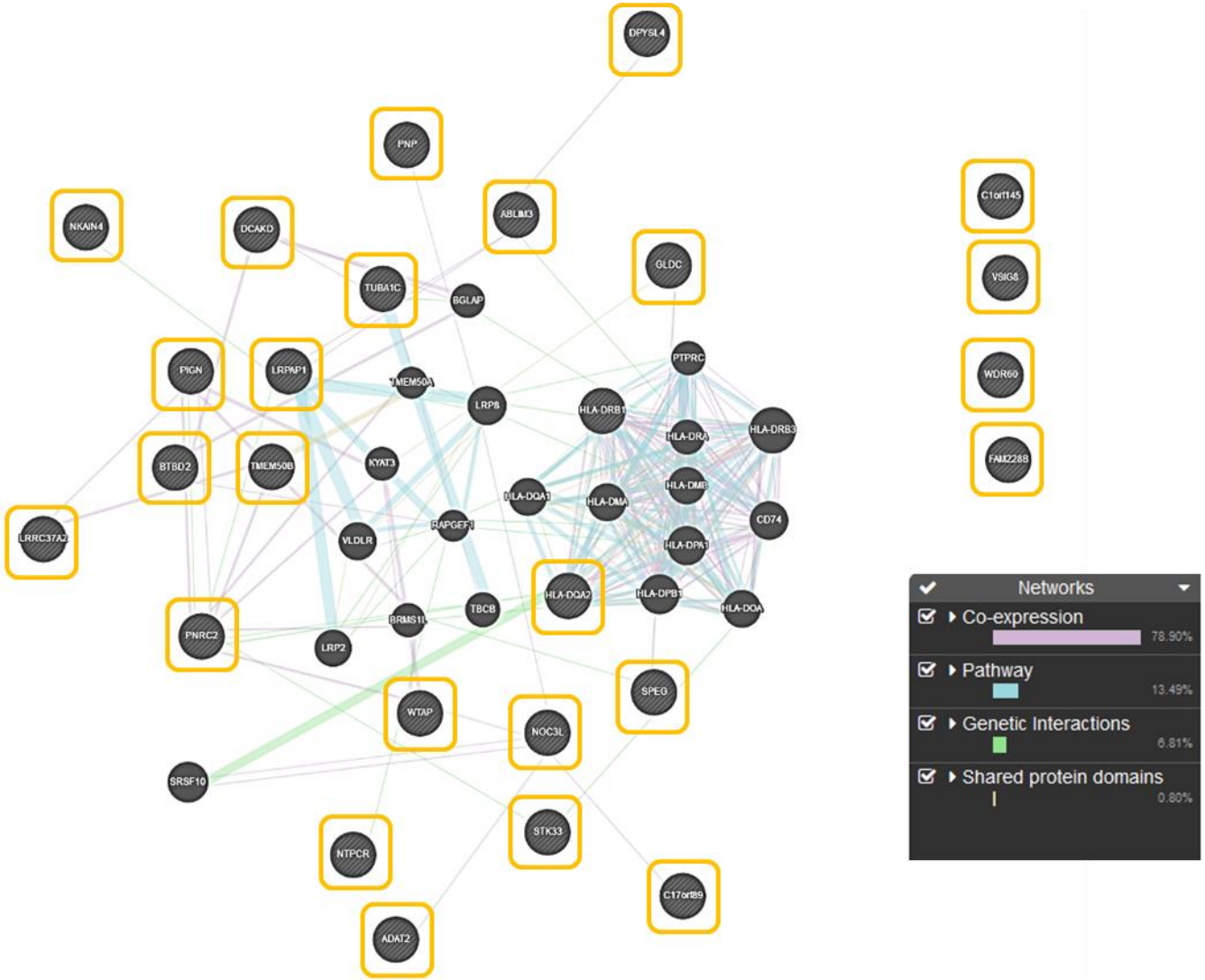


Figure 3: ***Genes identified in the comparison between AD and prostate cancer****. The network identifies different relationships shared between the genes in terms of co-expression (purple lines), pathway (blue lines) and genetic interactions (green lines).*

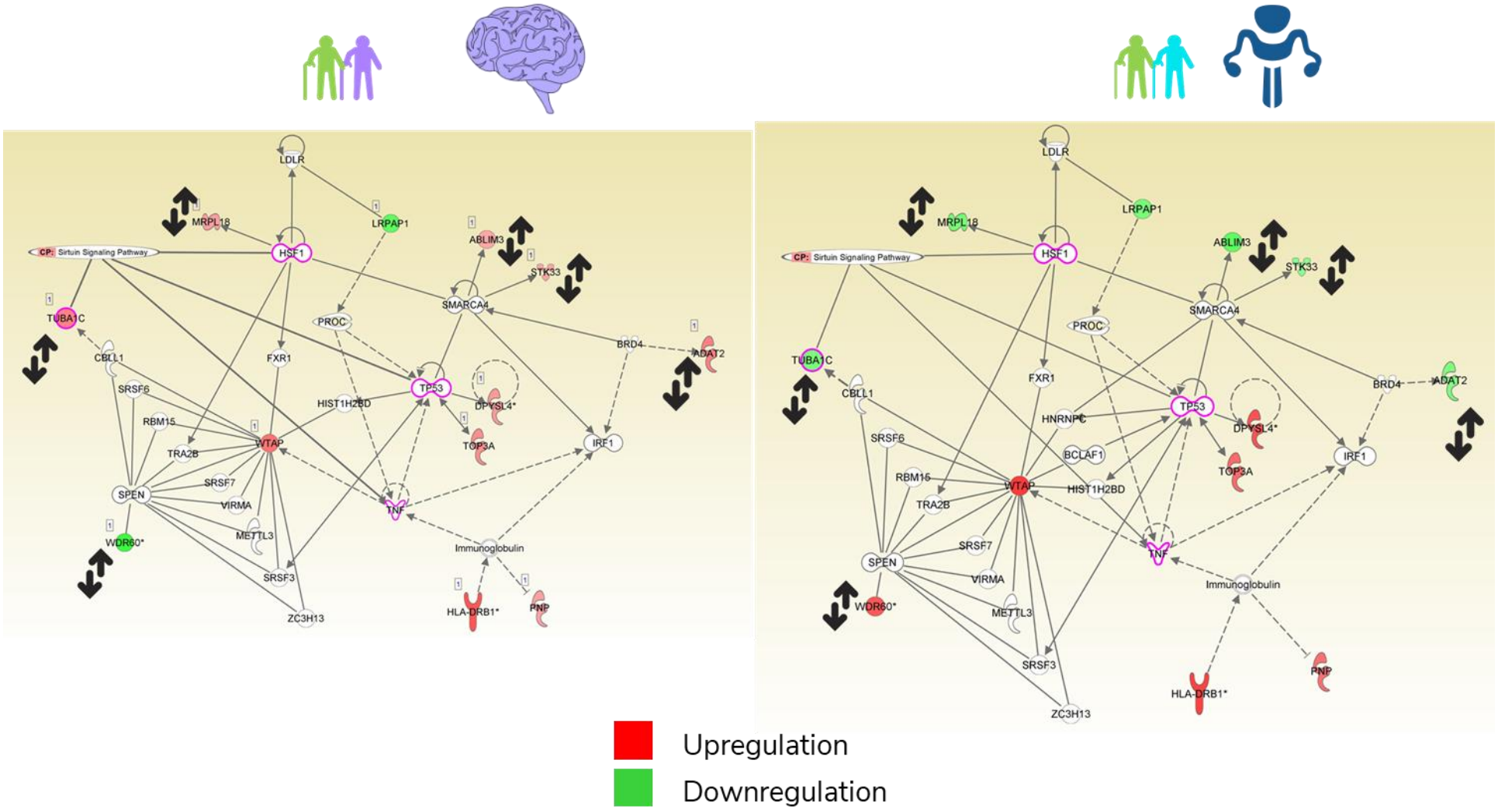


*Figure 4: **Enriched pathways using IPA's biobase**. The network connecting the genes associated between AD and prostate cancer indicates involvement of sirtuin signaling. Both panel represent the same pathway; left panel shows the gene expression of control vs AD association in brain tissues, and the right panel shows the gene expression of control vs cancer in prostate tissue.*

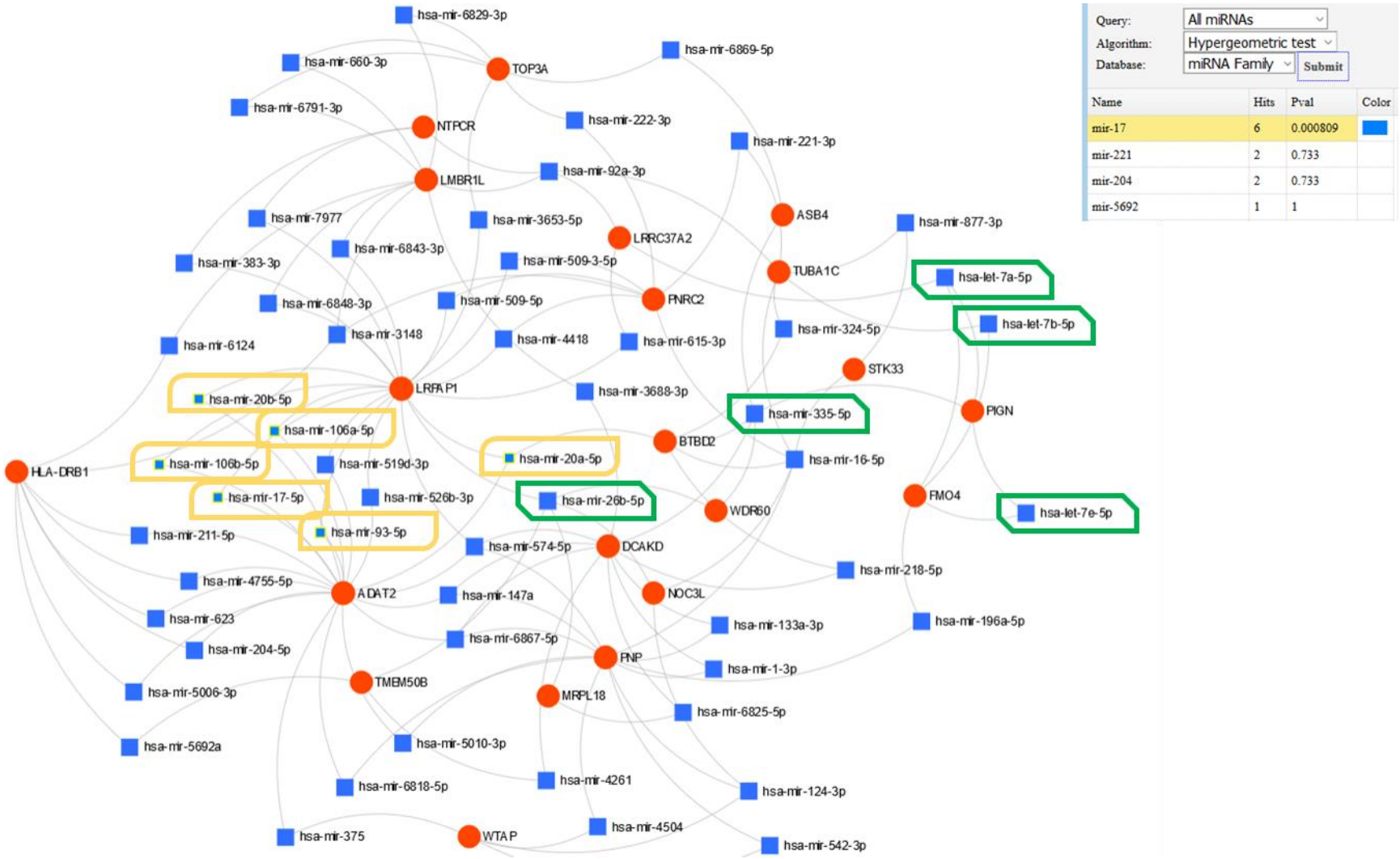


Figure 5: ***miRNA network for genes identified between AD and prostate cancer****. The miRNAs (blue squares) are mapped to genes (red circles), followed by testing of miRNA family enrichment (top panel). The members of the significant miRNA family – miR-17 is highlighted in yellow. The miRNAs highlighted in green are miRNAs along with miRNA-17 that we found consistent with our previous finding as well.*

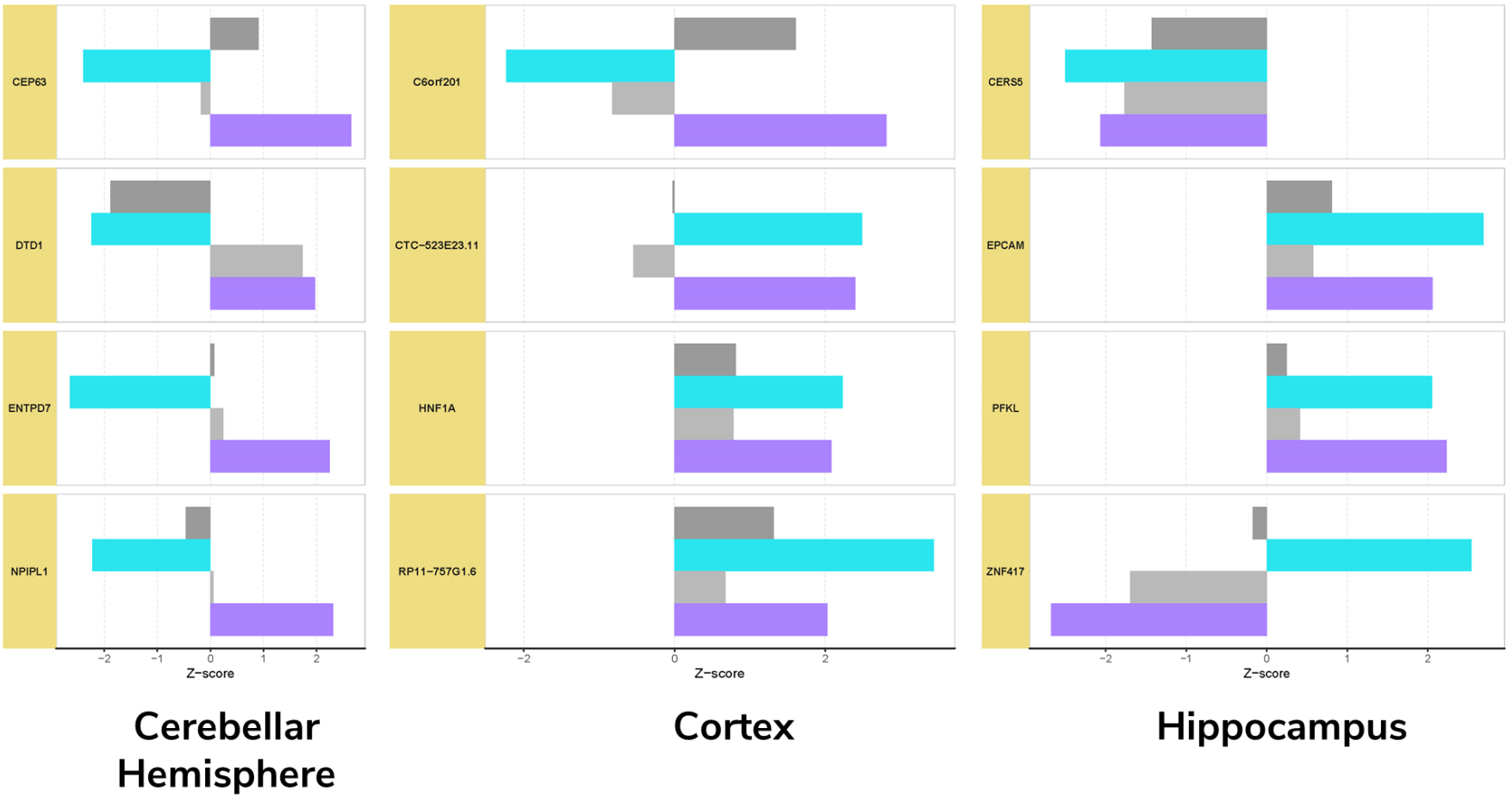
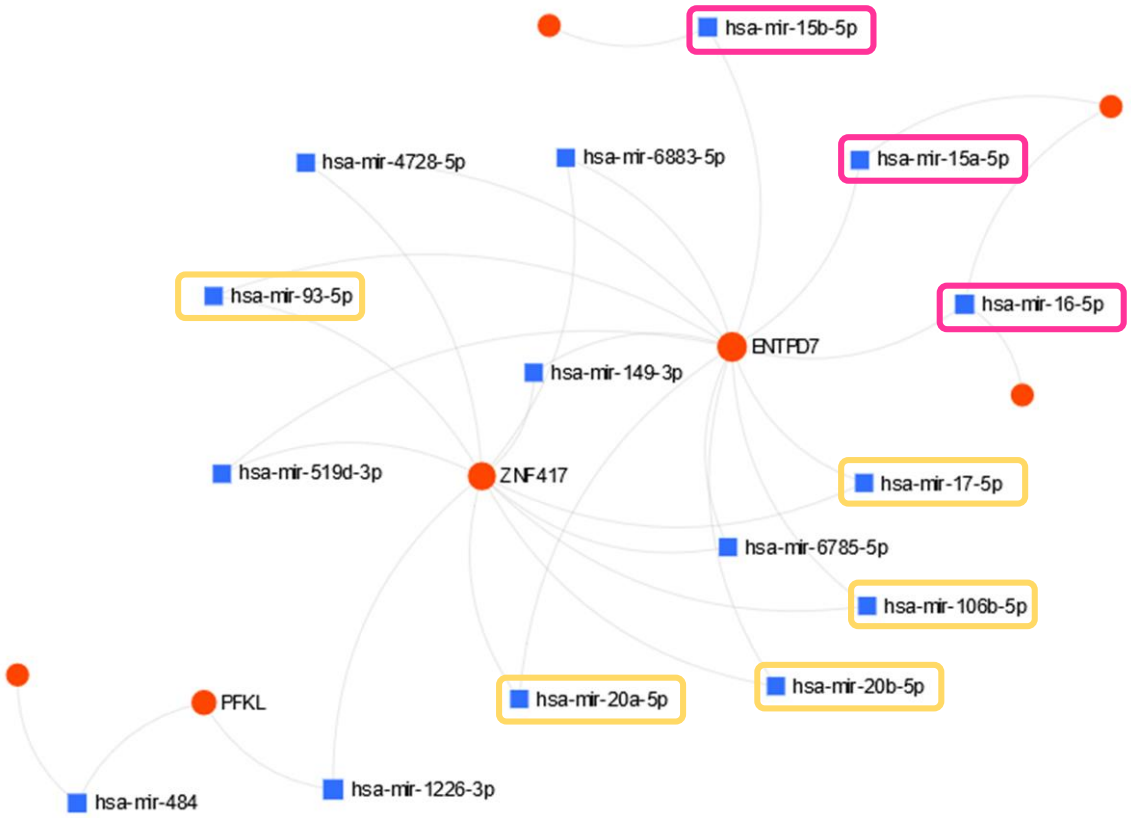


Figure 6: ***Transcriptomic profile of AD and breast cancer.*** *The genes that were associated with AD in brain tissues and with breast cancer in breast tissue are listed here, derived from each pairwise comparison of brain and cancer tissue. The z-score identify if the gene is up/downregulated. Blue bars show significant association for ctrl vs cancer in females and purple bars show significant association for ctrl vs AD females. Grey bars represent that these genes were not significant between ctrl vs cancers in brain tissue and ctrl vs AD were not significant in breast tissue.*



Query: All miRNAs
Algorithm: Hypergeometric test
Database: miRNA Family Submit

Name	Hits	Pval	Color
mir-17	5	0.0000292	
mir-15	3	0.0125	
mir-1226	1	1	
mir-515	1	1	
mir-484	1	1	
mir-149	1	1	

Figure 7: ***miRNA network for genes identified between AD and breast cancer****. The miRNAs (blue squares) are mapped to genes (red circles), followed by testing of miRNA family enrichment (top panel). The members of the significant miRNA family – miR-17 is highlighted in yellow and pink rectangles highlight miRNA-15 family members.*

5.5 SUPPLEMENTARY FILES

Link to AD vs prostate cancer - Males-ADvsCancer

Link to AD vs breast cancer - Females- ADvsCancer

5.6 REFERENCES

1 Franceschi, C. *et al.* The Continuum of Aging and Age-Related Diseases: Common Mechanisms but Different Rates. *Front Med (Lausanne)* **5**, 61-61, doi:10.3389/fmed.2018.00061 (2018).

2 Bonda, D. J. *et al.* The sirtuin pathway in aging and Alzheimer disease: mechanistic and therapeutic considerations. *Lancet Neurol* **10**, 275-279, doi:10.1016/S1474-4422(11)70013-8 (2011).

3 Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* **66**, 271-289, doi:10.3322/caac.21349 (2016).

4 Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* **8**, 193-204, doi:10.1038/nrc2342 (2008).

5 Aunan, J. R., Watson, M. M., Hagland, H. R. & Soreide, K. Molecular and biological hallmarks of aging. *The British journal of surgery* **103**, e29-46, doi:10.1002/bjs.10053 (2016).

6 Feng, Y. A. *et al.* Investigating the genetic relationship between Alzheimer's disease and cancer using GWAS summary statistics. *Hum Genet* **136**, 1341-1351, doi:10.1007/s00439-017-1831-6 (2017).

7 Kong, W. *et al.* Differences of immune disorders between Alzheimer's disease and breast cancer based on transcriptional regulation. *PloS one* **12**, e0180337-e0180337, doi:10.1371/journal.pone.0180337 (2017).

8 Frain, L. *et al.* Association of cancer and Alzheimer's disease risk in a national cohort of veterans. *Alzheimers Dement* **13**, 1364-1370, doi:10.1016/j.jalz.2017.04.012 (2017).

9 Ibanez, K., Boullosa, C., Tabares-Seisdedos, R., Baudot, A. & Valencia, A. Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. *PLoS Genet* **10**, e1004173, doi:10.1371/journal.pgen.1004173 (2014).

10 Tabares-Seisdedos, R. & Rubenstein, J. L. Inverse cancer comorbidity: a serendipitous opportunity to gain insight into CNS disorders. *Nat Rev Neurosci* **14**, 293-304, doi:10.1038/nrn3464 (2013).

11 Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091-1098, doi:10.1038/ng.3367 (2015).

12 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

13 Cacace, R., Sleegers, K. & Van Broeckhoven, C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimers Dement* **12**, 733-748, doi:10.1016/j.jalz.2016.01.012 (2016).
14 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).
15 Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat Protoc* **5**, 1564-1573, doi:10.1038/nprot.2010.116 (2010).
16 Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research* **38**, W214-W220, doi:10.1093/nar/gkq537 (2010).
17 Ribeiro, P. *et al.* miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic Acids Research* **44**, W135-W141, doi:10.1093/nar/gkw288 %J Nucleic Acids Research (2016).
18 Carafa, V., Altucci, L. & Nebbioso, A. Dual Tumor Suppressor and Tumor Promoter Action of Sirtuins in Determining Malignant Phenotype. *Frontiers in pharmacology* **10**, 38-38, doi:10.3389/fphar.2019.00038 (2019).
19 Podtelezhnikov, A. A. *et al.* Molecular insights into the pathogenesis of Alzheimer's disease and its relationship to normal aging. *PloS one* **6**, e29610-e29610, doi:10.1371/journal.pone.0029610 (2011).
20 Cacabelos, R. *et al.* Sirtuins in Alzheimer's Disease: SIRT2-Related GenoPhenotypes and Implications for PharmacoEpiGenetics. *International journal of molecular sciences* **20**, 1249, doi:10.3390/ijms20051249 (2019).
21 Chalkiadaki, A. & Guarente, L. The multifaceted functions of sirtuins in cancer. *Nature Reviews Cancer* **15**, 608, doi:10.1038/nrc3985 (2015).
22 Lalla, R. & Donmez, G. The role of sirtuins in Alzheimer's disease. *Frontiers in aging neuroscience* **5**, 16-16, doi:10.3389/fnagi.2013.00016 (2013).
23 Fujita, Y. & Yamashita, T. Sirtuins in Neuroendocrine Regulation and Neurological Diseases. *Front Neurosci* **12**, 778-778, doi:10.3389/fnins.2018.00778 (2018).
24 Fourcade, S., Outeiro, T. F. & Pujol, A. SIRT2 in age-related neurodegenerative disorders. *Aging (Albany NY)* **10**, 295-296, doi:10.18632/aging.101397 (2018).
25 Tanzer, A. & Stadler, P. F. Molecular evolution of a microRNA cluster. *Journal of molecular biology* **339**, 327-335, doi:10.1016/j.jmb.2004.03.065 (2004).
26 Huang, E., Liu, R. & Chu, Y. miRNA-15a/16: as tumor suppressors and more. *Future Oncology* **11**, 2351+ (2015).
27 Lovat, F. *et al.* miR-15b/16-2 deletion promotes B-cell malignancies. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 11636-11641, doi:10.1073/pnas.1514954112 (2015).
28 Gong, G. *et al.* miR-15b represses BACE1 expression in sporadic Alzheimer's disease. *Oncotarget* **8**, 91551-91557, doi:10.18632/oncotarget.21177 (2017).
29 Delay, C., Calon, F., Mathews, P. & Hébert, S. S. Alzheimer-specific variants in the 3'UTR of Amyloid precursor protein affect microRNA function. *Molecular neurodegeneration* **6**, 70-70, doi:10.1186/1750-1326-6-70 (2011).
30 Liu, F. *et al.* Prognostic role of miR-17-92 family in human cancers: evaluation of multiple prognostic outcomes. *Oncotarget* **8**, 69125-69138, doi:10.18632/oncotarget.19096 (2017).
31 Tong, Z. *et al.* VDAC1 deacetylation is involved in the protective effects of resveratrol against mitochondria-mediated apoptosis in cardiomyocytes subjected to

anoxia/reoxygenation injury. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **95**, 77-83, doi:10.1016/j.biopha.2017.08.046 (2017).

Chapter 6

DISCUSSION & FUTURE DIRECTIONS

6.1 DISCUSSION

Aging is a gradual decline of physiological process in the organism. Unfortunately, this gradual decline can manifest in different pathologies based on genetic and lifestyle factors. Epidemiological observations have reported parallel correlation of two or more disease pathologies with increase in ages of 65 years and older.

In this study, we investigated comorbidity patterns present within four aging associated diseases – radiotherapy side-effects, cancer, Alzheimer's disease and hypertension. To understand comorbidity patterns we employed integration of multidimensional genetic data for each disease comorbidity under investigation.

With the improvement of cancer therapies, survival rates of most prevalent cancers such as prostate cancer has increased to 90% for a period of 5-years. However, even the most precise dose and cancer-site radiation therapy can result in long-term adverse effects. We investigated radiation-induced proctitis which is a inflammation of the rectum and prevalent in 5-20% of cancer survivors. We integrated transcriptomics derived from genotype data, and copy number association to find possible genetic links with proctitis in prostate cancer survivors. The SNP-derived transcriptomic profiles identified novel genes that are co-expressed with known DNA-damage

response genes in prostate and whole-blood tissue. We integrated tissue specific protein-protein interaction information to identify genes that possibly play role in radiotherapy adversity. The goal of the study is to pave the way for translational application of genetic variants to identify individuals who may be at higher susceptibility of developing proctitis from radiation therapy, thus improving quality of life for cancer survivors. The role of genetics in radiotherapy is fairly new, and will benefit from recruitment of larger cohorts in multiple cancer types. This study provided insight into prostate cancer radiation therapy side effects, which may be different from other cancer types.

Next, we compared the direct comorbidity pattern of Alzheimer's disease and hypertension. Although the role of hypertension as metabolic risk factor to AD has been investigated by several studies, genetic differences between AD individuals with and without hypertension remains understudied. Due to hypertension's role in causing vascular dementia, we investigated several known clinical traits of vascular dementia in individuals with AD-hypertension comorbidity. It was interesting to note that hypertension was not correlated with white matter hyperintensity, indicating that the differences in genome-wide profile we reported in the study are likely due to presence of hypertension. We identified genes that play roles in cognitive impairement and hypertension seprately but have not been associated with AD-hypertension pathology. However, it is certainly possible that other metabolic factors might be associated with the differences we observed. For example, we observed that metabolic traits – cholestrol and triglycerides were correlated and triglyceride was correlated with systolic pressure, MMSE and age. Therefore, future studies are required to understand the role of these metabolic profiles and their effect on causation and progression of cognitive decline in AD patients. Furthermore, our study identified sex-specific differences in gene expression of candidate hits. The prevalence of AD-hypertension comorbidity

in males and females along with deep phenotyping will provide much needed resolution in the associated risk factors.

We also investigated indirect comorbidity shared between AD and cancer. This relationship of inverse correlation between two pathogenically different diseases faces continued skepticism in the scientific community. However, as reported earlier, we chose a conservative approach under the Bayesian umbrella to investigate this 'suspicious' relationship between AD, and breast and prostate cancer. Several literature and meta-analysis studies have hypothesized pathways that may be inverse between AD and cancer. By evaluating the role of each variant towards either disease, allowed to identify targets that have opposite proportions in AD and cancer. Even though we found few targets, the literature supported our reported findings. As we continued to remain cautious of our findings, we explored other avenues of mechanism that could be driving the effect we reported between AD and cancer. We conducted a comprehensive literature search on the mapped miRNAs to investigate their role with our targeted regions and their association with AD and cancer. We believe that the SNPs reported here have understudied allies that alter gene expression via miRNA regulatory mechanism. This study has major potential for translational impact, as the targets we reported here are protective in one disease and risky in another. These findings encourage investigating mutiallelic SNPs in the context of variants which play dual role in therapeutically-challenging diseases.

It is often expected that transcriptomic and gene variant profiles rarely overlap for multifactorial diseases. In our extended study of SNP-derived tissue-specific gene expression profile, we observed a similar trend. Even though we predicted gene expression from the same genotype profile, we identified different set of associated genes in AD and cancer. Here, we observed another interesting possibility of sirtuin signaling playing a role between AD and cancer. Combining the

findings from both studies, we believe that mitochondrial dysfunction due to alterations in sirtuin protein with *TOMM40* and *VDAC1* may be responsible for contrasting pathologies in the two diseases. Another role of these genes is the difference in contributing type-2 and type-3 epithelial-mesenchymal-transition in AD and cancer, which has not been investigated thoroughly.

Upon stratification by severity of disease or multiple conditions, the studies conducted here lose power to detect all possible genes that play role in comorbidity being investigated. All our studies were conducted in individuals of European ethnicity, due to availability of cohorts with required genetic and clinical information. It is imperative that these methods of investigation be applied in other ethnicities to identify underlying genetic associations of comorbidity patterns. In addition to the proposed studies, mild cognitive impairement which is a AD precursor was investigated in Mexican Americans by analysing differences in their epigenomic profiles (Appendix).

6.2 LIMITATIONS

The designated integrative genomics approach using prior biological knowledge to reconstruct networks is one of the most intuitive methods, giving appropriate context for data interpretation. This approach does have its limitations, however; because it is based upon existing network knowledge, it does not identify de-novo relationships. This study endeavors to investigate multiple facets of contributing factors to complex disease, however, other factors beyond the scope of the study may be affecting the disease as well. Rare variant effects may be missed as the study population is smaller than 10,000.

6.3 FUTURE STUDIES

This research study could further benefit from integration of methylation data. The current study primarily investigates age-associated disease and their comorbidity patterns in Caucasian ethnicity

and has the potential to be extended to other ethnicities to develop strategies for better tailored clinical decisions and treatments. Future work may include evaluating the impact of hypertension in chemotherapy treated cancer survivors, to determine genetic factors that may contribute to the development of the hypertension due to cancer treatment. Furthermore, integrative genomics approach with deep phenotyping to would be valuable to comprehend multimorbidity - presence of more than two chronic conditions. Studying the influence of SNPs on network of biology of complex disease will shorten the translational window to achieve personalized treatments.

Finally, this study underscores genetic heterogeneity that exists in comorbid condition, which are highly prevalent in the aging population. In order to provide precision diagnosis and treatment to the growing elderly population, it is imperative that more studies are conducted within the scope of multiple diseases.

APPENDIX

Genome-wide methylation of mild cognitive impairment in Mexican Americans highlights genes involved in synaptic transport, AD-precursor phenotypes, and metabolic morbidities

Gita A Pathak[1], Talisa K Silzer[1], Jie Sun[1], Zhengyang Zhou[2], Ann A Daniel[1], Leigh Johnson[3,4], Sid O'Bryant[3,4], Nicole R Phillips[1], Robert C Barber[4] *

ABSTRACT

The Mexican American population is among the fastest growing aging population and has a younger onset of cognitive decline. This group is also heavily burdened with metabolic conditions such as hypertension, diabetes and obesity. Unfortunately, limited research has been conducted in this group. Understanding methylation alterations, which are influenced by both genetic and lifestyle factors, is key to identifying and addressing the root cause for mild cognitive impairment, a clinical precursor for dementia. We conducted an epigenome-wide association study on a community-based Mexican American population using the Illumina EPIC array. Following rigorous quality control measures, we identified 10 CpG sites to be differentially methylated between normal controls and individuals with mild cognitive impairment annotated to *PKIB, KLHL29, SEPT9, OR2C3, CPLX3, BCL2L2-PABPN1* and *CCNY*. We found four regions to be differentially methylated in *TMEM232, SLC17A8, ALOX12* and *SEPT8*. Functional gene-set analysis identified four gene-sets - *RIN3, SPEG, CTSG and UBE2L3* as significant. The gene ontology and pathway analyses point to neuronal cell death, metabolic dysfunction and inflammatory processes. We found 1450 processes to be enriched using empirical Bayes gene-set enrichment. In conclusion, the functional overlap of differentially methylated genes associated with cognitive impairment in Mexican Americans implies cross-talk between metabolically-instigated systemic inflammation and disruption of synaptic vesicular transport.

Keywords: Alzheimer's Disease, Epigenetics, Mexican Americans, SEPT8 protein, Cognitive Dysfunction

Introduction

Alzheimer's disease (AD) affects 5.8 million Americans and is currently the 6th most common cause of death (1). AD is a neurodegenerative disease that manifests as a result of intracellular tau tangles and extracellular amyloid beta plaques within the brain, causing cellular inflammation and neuronal loss, which lead to changes in memory and personality (1). Mild cognitive impairment (MCI), a common precursor to AD, affects over a quarter of individuals over the age of 65, with a third of these individuals developing AD later in life (2).

Within the United States, the Mexican American elderly population is experiencing rapid growth and is predicted to increase 6-fold by 2050 (3). This population has an earlier age of AD onset (4) and are typically diagnosed at more advanced stages of disease relative to non-Hispanic whites (3, 4). Further, Mexican Americans are disproportionately burdened with comorbid conditions such as depression (5, 6), cardiovascular disease (7) and metabolic conditions (e.g. metabolic syndrome, type 2 diabetes, obesity) (4, 7). Other factors such as education and socio-economic status have also been suggested to play a role, where fewer years of education and lower economic status are known to increase one's risk for AD (8).

There are several critical differences in the etiology of AD when comparing Mexican Americans and non-Hispanic whites, and these distinctions may be key to understanding the observed disparities in the population. Mexican Americans are less likely to carry the APOE e4 allele which is the one of the highest predisposing genetic factors for AD among non-Hispanic whites (4, 8). Additionally, while a vascular/inflammatory phenotype predominates risk for cognitive decline

among non-Hispanic whites, risk among Mexican Americans is due primarily to metabolic factors (9). In a study of serum biomarkers from 363 Mexican Americans (AD=49, NC=314), the biomarker profile of AD among Mexican Americans was weighted heavily for metabolic markers as opposed to the inflammatory-laden profile that has previously been observed among non-Hispanic whites (10).

Cognitive impairment is highly heterogeneous, with both environmental/lifestyle and genetic factors playing a role. Since AD pathology can initiate up to 20 years before the first visible symptoms (1), changes in methylation patterns may signify relevant early molecular responses to the disease. Methylation at CpG (cytosine-phosphate-guanine) sites throughout the genome is known to affect downstream expression of various genes and has been implicated in normal aging and disease processes (11, 12). Differential methylation relating to AD and MCI has already been reported in several populations (11-14). While metabolic stress detected in the peripheral blood is associated with cognitive impairment, the exact mechanisms are unknown (15). Wang et al. have identified epigenetic changes in the brain of AD patients that exacerbate synaptic plasticity within functional networks overlapping multiple brain regions (16). Moreover, changes in DNA methylation patterns have been shown to be highly correlated between peripheral blood and brain (17). Since metabolic burden affects both the peripheral and central-nervous system (18), detecting epigenetic signatures may help identify risk regions that overlap metabolic burden and cognitive dysfunction.

Here we sought to investigate genome-wide differential methylation at site- and region-specific levels to identify epigenetic factors that may confer risk for cognitive impairment in Mexican Americans.

Results

Genome-wide methylated sites and regions

In the genome-wide probe analysis, 10 CpG probes were found to be significantly (FDR p-value < 0.05) differentially methylated probes (DMPs) between normal controls (NC) and individuals with mild cognitive impairment (MCI) (Supplementary file 1). These CpG sites were located within *PKIB, KLHL29, SEPT9, OR2C3, CPLX3, PABPN1-BCL2L* and *CCNY* (Figure 1).

The genome-wide region analysis yielded 4 differentially methylated regions (DMRs): two on chromosome 5 located within *TMEM232* and *SEPT8*, one on chromosome 12 within *SLC17A8*, and one on chromosome 17 within *ALOX12* (Supplementary file 2). These 4 regions comprised 41 significant sites that were located primarily within CpG islands, with some being in open sea and shore regions (Figure 2).

Differentially methylated modules

The functional methylation analysis identified four gene-sets (modules) to be significantly hypomethylated in MCI compared to NC – *RIN3* (also known as *SLC24A4*), *SPEG, CTSG and UBE2L3* (Supplementary file 3). The *RIN3 and SPEG* cluster identified overlapping significant genes - *RIN3, SPEG, XDH* and *KNDC1*. While in the *CTSG* cluster we found *CTSG, F2RL1* and *MMRN1* to be significant. Lastly, in the *UBE2L3* gene-set, we found *RNF144A*, *UBE2L3*, and *NEDD4L* to be significant (Figure 3).

Gene Ontology and pathway enrichment test

Differentially methylated genes identified through DMPs, DMRs, and functional network analysis were screened for enrichment based on gene ontology (Supplementary file 4). Several biological processes were found to be significantly enriched, which clustered primarily on immunological processes, and systemic arterial blood pressure (Figure 4). Four KEGG pathways – synaptic vesicle

cycle, bacterial invasion of epithelial cells, inflammatory mediator regulation of TRP, and ubiquitin mediated proteolysis were found to be enriched. Using IPA (Ingenuity Pathway Analysis®) to assess enriched disease pathways, we found neuronal cell death and metabolic disease (insulin-resistance) among the most significant pathways (Figure 5).

Bayes Gene-Set Enrichment Analysis

We also conducted another gene-set enrichment analysis using the empirical Bayes method (19), which ranks genes into gene-sets using beta-values for each of the CpG sites. We found a total of 1450 gene-sets to be significantly enriched (adjusted P-value < 0.05). With respect to the most significantly-enriched, many of the gene-sets have been associated with H3K27 methylation in various experimental models (Table 1 in Supplementary file 5), while those with the highest classification accuracy (~72%-78% of the area under the curve (AUC)) were attributed to the following gene-sets (among others, Table 2 in Supplementary file 5): *Regulation of protein*, *Rho GTPase activity*, and *Aging*.

Discussion

The etiology of MCI/AD in Mexicans- Americans is distinctive from non-Hispanic whites. Much of this disparity has been attributed to the elevated prevalence of metabolic dysfunction within this population, with 43% of individuals with dementia also suffering from diabetes, stroke or both (8). While several studies have reported an association between metabolic burden and cognitive impairment, the exact mechanisms linking the two are poorly understood (20). In the last 20 years, based on pubmed query, we found only 71 articles that were published in the genetics of Alzheimer's in Mexican-American population. There are no publications listed when we replace the term genomics and its associated terms with 'methylation' or 'epigenetics' (Supplementary

figure 2). Studying alterations in genome-wide methylation profiles associated with cognitive decline in a metabolically burdened admixed population, may help identify changes that occur as a result of both genetic and lifestyle factors. Understanding these factors will be vital for development of effective diagnostic and therapeutic strategies.

Genes implicated in differentially methylated sites and regions

Here we found 10 differentially methylated CpG sites significantly associated with MCI in our Mexican American cohort. Four of these sites were hypomethylated among participants with mild cognitive impairment than normal controls; cg25016219 (*KLHL29*), cg26479998 (*SEPT9*), cg02586267 (not mapped to a gene), cg18978297 (*CPLX3*). *KLHL29* has been reported to be associated with diabetes (21), while other members of the Kelch-like gene family have been implicated in hypertension (22). Septin 9 (*SEPT9*), known to be involved in cell cycle regulation, has primarily been associated with neuralgic amyotrophy, a peripheral nervous system disorder characterized by weakness and atrophy of the upper extremities (23). *SEPT9* variants have also been associated with systolic blood pressure (24). The hypomethylation of *SEPT9* observed in our analysis was also reported by Dayeh and colleagues, who found hypomethylation of *SEPT9* on multiple CpG sites and an upregulation of *SEPT9* in human pancreatic islets from type 2 diabetes donors (25). Complexin 3 (*CPLX3*) is involved in regulation of SNARE-associated synaptic vesicle fusion, thereby impacting neurotransmitter transporter activity and localization (26). The CpG site within *CPLX3* was found to be hypomethylated in our study based on peripheral blood, and Annese et. al have reported *CPLX3* to be upregulated in the hippocampal region of individuals with late-onset Alzheimer's disease (27).

Six CpG sites – cg22360048 (*PKIB*), cg20904111 (intergenic), cg05917713 (*BCL2L2-PABPN1*), cg20201669 (*OR2C3*), cg14179796 (*CCNY*) and cg22327037 (intergenic) were found to be

hypermethylated in individuals with MCI. PKIB is a kinase inhibitor in PKA signaling known to suppress insulin secretion under chronic hyperglycemia (28). Interestingly, PKA signaling also regulates mitochondrial function/morphology as well as neuronal survival, and has been suggested as a therapeutic target for treatment of neurodegenerative diseases (29). *BCL2L2* (*BCL-w* or *BCL2-w*), encoding an anti-apoptotic protein, was hypermethylated in our study, and has been reported to be downregulated in AD brains afflicted with increased phosphorylation of tau proteins (30). Furthermore, variants in *BCL2L2-PABPN1* have been reported to be associated with basal metabolic rate and body mass index in obese Korean women (31). Mutations in *PABPN1* encoding a poly A binding protein, have been associated with increased protein aggregation, a pathological hallmark of many neurodegenerative diseases (32). Olfactory deficits have been reported to precede clinical symptoms of AD and MCI (33), and our study identified hypermethylation of *OR2C3*, encoding a GPCR that binds to the extracellular domain that generates signal to the olfactory sensory neurons in the brain (34). *OR2C3* binding results in additional activation of Akt, MAPK and Rho-signaling pathways (34). Therefore, it is plausible that hypermethylation of sites within *OR2C3* may cause downregulation of its expression resulting in olfactory deficits and neuronal dysregulation (35). The CpG site within Cyclin Y (*CCNY*) was observed to be hypermethylated in our study; downregulation of *CCNY* has been reported to inhibit the formation of new synapses during synaptic remodeling (36). *CCNY* also regulates Wnt-signaling pathways, which have been found to be altered in the prefrontal cortex of individuals with Alzheimer's disease (37)

We also identified four regions within *TMEM232*, *SEPT8*, *SLC17A8*, and *ALOX12* that were significantly differentially methylated. *ALOX12* - arachidonate 12-oxidoreductase, is an enzyme involved in the generation of lipid metabolites and loss of function mutants of *ALOX12* (mouse

model) have demonstrated sensitization of pancreatic beta cells to oxidative stress (38) leading to declines in insulin secretion, an event that precedes onset of type 2 diabetes (39). Interestingly, *ALOX12* has also been reported to modulate glutamate-induced degeneration of neurons (40), and SNPs within this gene have been implicated in loss of prefrontal cortical thickness in PTSD individuals (41). Vesicular glutamate transporter 3 (otherwise known as *VGLUT3* or *SLC17A8*) has been shown to be co-expressed with glutamate receptor GABRG3 in peripheral glutamatergic neurons; together they are suspected to play a compensatory role in response to loss of glycinergic activity which occurs during hearing loss (42), a precursor symptom to MCI and AD (43). Another member of the Septin gene family, *SEPT8*, was found here to be associated with MCI. *SEPT8* is believed to regulate the transportation of synaptic vesicles through interaction with different SNARE- complex components (44). Dysfunction of synaptic vesicle transmission in hippocampal and cortical regions has been associated with cognitive decline (45). Additionally, multiple variants in *TMEM232*, encoding a transmembrane protein, have been implicated in various neuropathies (46). SNPs in this gene have also been associated with schizophrenia, depression and bipolar disorder in a Chinese population(47).

Genes implicated in differential methylated modules

We found four gene-sets in our study – (a) *RIN3*, (b) *SPEG*, (c) *CTSG* and (d) *UBE2L3*, to be significantly hypomethylated in individuals with MCI. Four genes- *RIN3*, *SPEG*, *XDH* and *KNDC1* were hypomethylated in both *RIN3* and *SPEG* gene-sets. Hypomethylation of CpG sites in the 3'UTR of *RIN3* has also been observed in peripheral blood of individuals with sporadic early onset AD (48) . Interestingly, in our study, the functional module analysis which weights location of CpG sites from TSS sites in an overall gene-based analysis, also identified hypomethylation of *RIN3* derived from peripheral blood of individuals with MCI, suggesting that hypomethylation at

this site may serve as an early indicator of cognitive decline. Due to its involvement in regulation of endocytic trafficking and processing of APP (49), hypomethylation of *RIN3* may also serve as a compensatory response to amyloid toxicity. This postulation is supported by Boden et al. who observed increased hypomethylation of *RIN3* in AD brains in comparison to blood, which they suggest to be due to prolonged exposure of the brain to an amyloid-rich environment (48). In a study by Stage et al. a single SNP within *RIN3* (rs10498633) was found to be associated with gray matter density within the medial temporal lobe, a phenotypic change detectable even in early stages of cognitive impairment (50). Other variants within *RIN3* have also been associated with metabolic syndrome in Japanese individuals (51). *KNDC1*, encodes a Ras-guanine nucleotide exchange factor, and Ji et al. have shown *KNDC1* upregulation to stimulate p53-oxidative stress responses (52). Other studies focusing on 5-hydroxymethylation in AD have found *KNDC1* to be enriched in GO terms related to neurogenesis and signal transduction (53). Remarkably, a GWAS of 1 million individuals, implicated variants in *KNDC1* to be associated with 'self-reported educational attainment', a highly correlated risk factor for cognitive impairment (54). Xanthine dehydrogenase (*XDH*), a hydroxylase involved in oxidative purine metabolism, is capable of converting to xanthine oxidase, a major producer of ROS (55). Elevated ROS is a hallmark of both AD and Type 2 diabetes (56). Upregulation of *XDH* activity has also been demonstrated in the brain of rats with late-stage diabetes (57). Hypomethylation of CpG sites within *XDH* may allow for increased expression permitting potential downstream elevation in ROS.

Cathepsin G (*CTSG*) is an anti-inflammatory protease found in azurophil granules of blood neutrophils that functions to breakdown pathogens and inflammatory tissues. Specifically, *CTSG* has been shown to play a role in abating neuroinflammation in AD through cleavage of amyloid beta (58). The hypomethylation of *CTSG* observed in our study, was reported to be upregulated

in the hippocampus of hypertensive aged rats suggesting perhaps that neuroinflammation may bridge hypertension and cognitive dysfunction (59). The G-protein coupled receptor *F2RL1* (otherwise known as *PAR2*), also accelerates neurodegeneration via inflammatory mediators (60). Interestingly, upregulation of *PAR2* has been reported to be induced by alpha-synuclein (61). *PAR2* has also been shown to be upregulated in adipose tissues and is a contributor to insulin resistance and metabolic dysfunction (62). While, the upregulation of *MMRN1* has been associated with cognitive deficits in Parkinson's disease(63).

The *UBE2L3* gene-set revealed several ubiquitin-related proteins to be hypomethylated. *UBE2L3* (*E2F1*) encodes a ubiquitin-conjugating enzyme, while *NEDD4L* and *RNF144A* encode E3 ubiquitin ligases. The expression of *UBE2L3* is upregulated in neurons of AD brains (64) and is also increased in obesity for maintenance of metabolic homeostasis (65). SNPs within *NEDD4L* have been associated with essential hypertension, a known comorbidity for AD (66). Hypomethylation and upregulation of mRNA expression of *RNF144* was reported in liver samples of individuals with type-2 diabetes (67). Further, the interactions between these enzymes are important for maintaining synaptic plasticity via selective degradation of abnormal or misfolded proteins (68).

Shared processes & pathways of differentially methylated genes based on gene-set enrichment

Three genes – *F2RL1, ALOX12* and *UBE2L3* were identified in both KEGG and IPA's knowledge base highlighting their role in inflammation, neuronal cell death and metabolic processes. We also conducted GSEA analysis using empirical Bayes approach which clusters differentially methylated probes into genes followed by overall ranking of genes based on differential methylation and testing for enrichment of gene-sets. Several experimentally-derived gene-sets

stored in MSigDB were found to be enriched in our genome-wide methylation analysis, with some of the top most gene-sets being: '*BENPORATH_ES_WITH_H3K27ME3*' (AUC – 60%; P-value : 5.1e-27), '*MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3*' (AUC – 59.3%; P-value: 7.5e-21), and '*BENPORATH_PRC2_TARGETS*' (AUC – 61%; P-value : 2.76e-19) . In order to understand their relevance to biological pathways, we tested gene members of these gene-sets for enriched pathways from the Panther database (www.pantherdb.org) (Supplementary file 5). We found four pathways to be common in all three gene-sets – (a) Alzheimer disease-presenilin pathway, (b) Ionotropic glutamate receptor pathway, (c) Heterotrimeric G-protein signaling pathway, and (d) Cadherin signaling pathway. (Figure 6)

Conclusion: Reported genes & pathways converge between cognitive dysfunction, metabolic burden and inflammation

Our results point to modification of genes involved in metabolic, neuroinflammatory and AD-precursor phenotypes (Figure 6). The data presented here links peripheral metabolic dysregulation with cognitive decline. Metabolic syndrome which includes obesity, diabetes and hypertension has been known to provoke poor cerebral blood flow, resulting in vasoconstriction and endothelial dysregulation; this leads to production of reactive oxygen species and inflammation (69). Insulin receptors are expressed in multiple brain regions, including the olfactory bulb, hypothalamus and hippocampus, and are sensitive to changes in the peripheral system due to metabolic conditions (70). Insulin resistance alters cAMP/PKA signaling pathways which not only induces inflammation from oxidative stress, but also affects synaptic plasticity of hippocampal neurons (71). In the presence of hypertension, homeostatic dysfunction results in microglia - induced inflammation in the brain (72). Metabolic syndrome – obesity, induces low-grade systemic inflammation which is detected by brain (73) causing disruption of the synaptic transporter activity

(74, 75). While our findings do not incriminate any single pathway for pathogenicity of cognitive dysfunction, they highlight genes and pathways involved in cross-talk between peripheral metabolic burden and neuroinflammation. We postulate that insults sustained from metabolic dysregulation results in low grade inflammation [reported earlier by our group] (10) affecting synaptic vesicle activity which ensues cognitive deficits in Mexican Americans. Therefore, it is possible that cognitive decline in Mexican-Americans is a manifestation of genetic/lifestyle factors of metabolic stress as seen here in their epigenetic profile.

There are several limitations to our study. First, our study is limited in sample size; therefore, the genes reported here should be considered as preliminary findings and should be replicated in larger cohorts. We have also tested cognitive-associated methylation changes in peripheral blood, which may not be parallel to brain-specific methylation alterations. Epigenetic variations in conjunction with RNA expression in larger cohorts of Hispanic individuals will provide much needed characterization of the pathogenesis of cognitive impairment.

Despite the limitations, we have strengthened our study design by implementing a very rigorous and conservative QC approach, correcting for multi-level batch effects, adjusting for cell composition and incorporating balanced demographic characteristics, representative of the Mexican American population. Our study has found several novel epigenetic markers associated with metabolic burden and cognitive dysfunction which contribute towards addressing the critical health disparity faced by Mexican Americans.

Methods & Materials

Samples and cohort design

This study was approved under the North Texas Regional IRB #2012–083. All participants provide written consent. Peripheral blood buffy coat of fasted-state Mexican American participants (n=90) enrolled in the Health & Aging Brains of Latino Elders (HABLE) study was obtained.

Each participant underwent an interview (i.e. medical history, medications, health behaviors), neuropsychological testing, blood draw, and medical examination. Global cognition was assessed via the Mini Mental State Examination (MMSE) (76) and disease severity rated according to the Clinical Dementia Rating scale (77) sum of boxes scores (CDR SB) (78, 79). An informant interview was conducted for each research participant to obtain information regarding his/her activities of daily living (basic and instrumental). All information was presented at a weekly consensus review conference with diagnoses of Alzheimer's disease (80) and MCI (81) assigned according to published criteria. Cognitively normal control (NC) participants performed within normal limits on psychometric assessment (82).

Phenotypic differences between the two groups (Normal Controls & Mild Cognitive Impairment) were tested for two-tailed significance using independent t-test for continuous variables, and chi-square test for categorical variables (Table 1). Participants were matched on sex and age. Peripheral blood samples were handled in accordance to the UNTHSC Institutional Biosafety Committee approved protocol IBC-2018-0078. Forty-five individuals (both male and female) diagnosed with mild cognitive impairment (based on a battery of neuropsychological tests) were selected. Forty-five age and sex-matched individuals were then chosen to complete the study population. Metabolic risk score (0-5) was calculated for each individual following adapted IDF guidelines based on obesity, elevated triglycerides, reduced HDL cholesterol, hypertension diagnosis and high fasting glucose (83). Waist circumference was used to diagnose obesity, with

35.4 inches being the cut-off value for Mexican American adults (84). Missing data was handled according to Masconi et al., 2015 (85).

Our study cohort is representative of the Mexican American population, which often a) is metabolically burdened (as reflected by high metabolic risk scores), b) has fewer years of education (on average) and, c) has low frequency of the APOE e4 allele. The most prevalent APOE status within our cohort was 3/3, with a single individual being 4/4; this was optimal for our study as the frequency and risk effect of the APOE e4 allele is often less in Hispanic populations (8).

Table 1. Demographic characteristics for the sample cohort.

	Normal Controls (NC) ***N= 45***	**Mild Cognitive Impairment (MCI)** ***N=45***	**P-value**
Mean ± SD			
Age	63.49 ± 7.20	63.53 ± 7.16	0.9790
MMSE	26.16 ± 2.94	23.87 ± 3.16	0.0006
Education	8.60 ± 4.35	6.11 ± 3.79	0.0048
Metabolic Score (IDF)	2.87 ± 1.41	3.36 ± 1.13	0.0723
N (%)			
Sex			
Female	32 (71%)	31(69%)	
Male	13 (29%)	14 (31%)	1.00
***APOE* Status**			
3/3	30 (67%)	29 (64%)	
3/4	12 (27%)	9 (20%)	
2/3	3 (7%)	6 (13%)	0.488
4/4	0 (0%)	1 (2%)	

DNA extraction and quantification

DNA was extracted from peripheral blood buffy coat using the MagBind® Blood and Tissue DNA HDQ Kit (Omega Bio-tek, Norcross, GA) on a Microlab STAR liquid handling system. DNA concentrations were quantified using Qubit® dsDNA BR Assay Kits with the Qubit® Fluorometer (Thermo Fisher Scientific Inc.) and the NanoDrop® spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). For downstream applications, DNA extracts with concentrations ≥10 ng/μL were considered sufficient due to the manufacturer's requirement of 200 ng of input DNA for both methylation and genotyping arrays. Low yield DNA extracts were concentrated using Microcon® DNA Fast Flow Filters (Sigma Aldrich, St. Louis, MO). Any extracts that remained <10 ng/mL following concentration, were excluded from downstream processing.

SNP genotyping

All 90 subjects were genotyped on the Infinium HTS Global Screening Array v.2 (Illumina, San Diego, CA) following manufacturer's instructions. Intensity data files were imported into Genome Studio®. All SNPs with call rates <0.99 were excluded from analysis. Remaining SNPs were QC'd in PLINK; SNPs with >5% missingness were removed (--geno) (86). The QC'd SNP dataset was used to identify presence of population substructure using ADMIXTURE (87). The cross-validation error rate was lowest for k of 1, concluding that the population did not have any sub-clusters (Supplementary Figure 1).

DNA methylation array assay

DNA methylation levels were analyzed for all 90 participants. DNA was bisulfite converted using the EZ DNA Methylation™ kit (Zymo Research, Irvine, CA). Bisulfite converted DNA was then processed on the Infinium HD MethylationEPIC Array (Illumina) following manufacturer's

guidelines. Intensity data (IDAT) files were imported into Genome Studio® and technical replicates were compared to determine consistency in array processing. Technical replicates all displayed R^2 values greater than the 0.98 suggested threshold (Illumina).

Data analysis and visualization

Raw IDAT files were processed using the 'ChAMP' Bioconductor package (88) in R, which converts fluorescent intensities into beta values (ratio of methylated:unmethylated) ranging from 0 to 1 for each individual probe. Data was normalized using the BMIQ method, followed by inspection for batch effects using Singular Value Decomposition (SVD). Both the lowest (i.e. sample plate) and highest (i.e. array) stages in the experimental workflow were corrected for using ComBat. Due to the consistent placement of technical replicates on the beadchips, one of each replicate pair was removed from analysis allowing for further correction of variance at the array level. The dataset for analysis included 738,919 CpG sites from 90 individuals. An association test to identify significantly differentially methylated probes was performed using M-values (transformed beta-values) and adjusted for variation in proportions of the following cell types – lymphocytes, monocytes and neutrophils, using limma (89). Differentially methylated probes were considered significant based on multiple testing corrected p-value < 0.05. Differentially methylated regions were analyzed using DMRcate(90). Functionally differentially methylated modules were also analyzed using normalized, batch-corrected beta-values (91) in ChAMP.

The significant genes identified were analyzed for gene ontology enrichment using empirical bayes GSEA (19) and the ShinyGO (92) tool set for biological processes, cellular components and molecular functions at an FDR < 0.05. Visualizations were created using Gviz (93) and ggplot2(94) in R. Candidate genes were analyzed for pathway enrichment, and subsequent figure

was generated through the use of IPA (QIAGEN Inc., https://www.qiagenbio-informatics.com/products/ingenuity-pathway-analysis software).

ACKNOWLEDGEMENTS

The research team thanks the local Fort Worth community and participants of the Health & Aging Brain Study. Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Numbers R01AG054073 and R56AG054073. We would also like to acknowledge the NIH – Neurobiology of Aging T32 grant AG020494 for supporting this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CONFLICT OF INTEREST STATEMENT

Dr. Sid O'Bryant has pending patents related to his Alzheimer's disease blood test. University of North Texas Health Science Center has licensed these patents to CX Precision Medicine, Inc. Dr. O'Bryant has a financial interest in this company, and Dr. O'Bryant is the Chief Scientific Advisor. CX Precision Medicine had no role in the design or results of this study.

Dr. Leigh Johnson has a financial interest in CX Precision Medicine. CX Precision Medicine had no role in the design or results of this study.

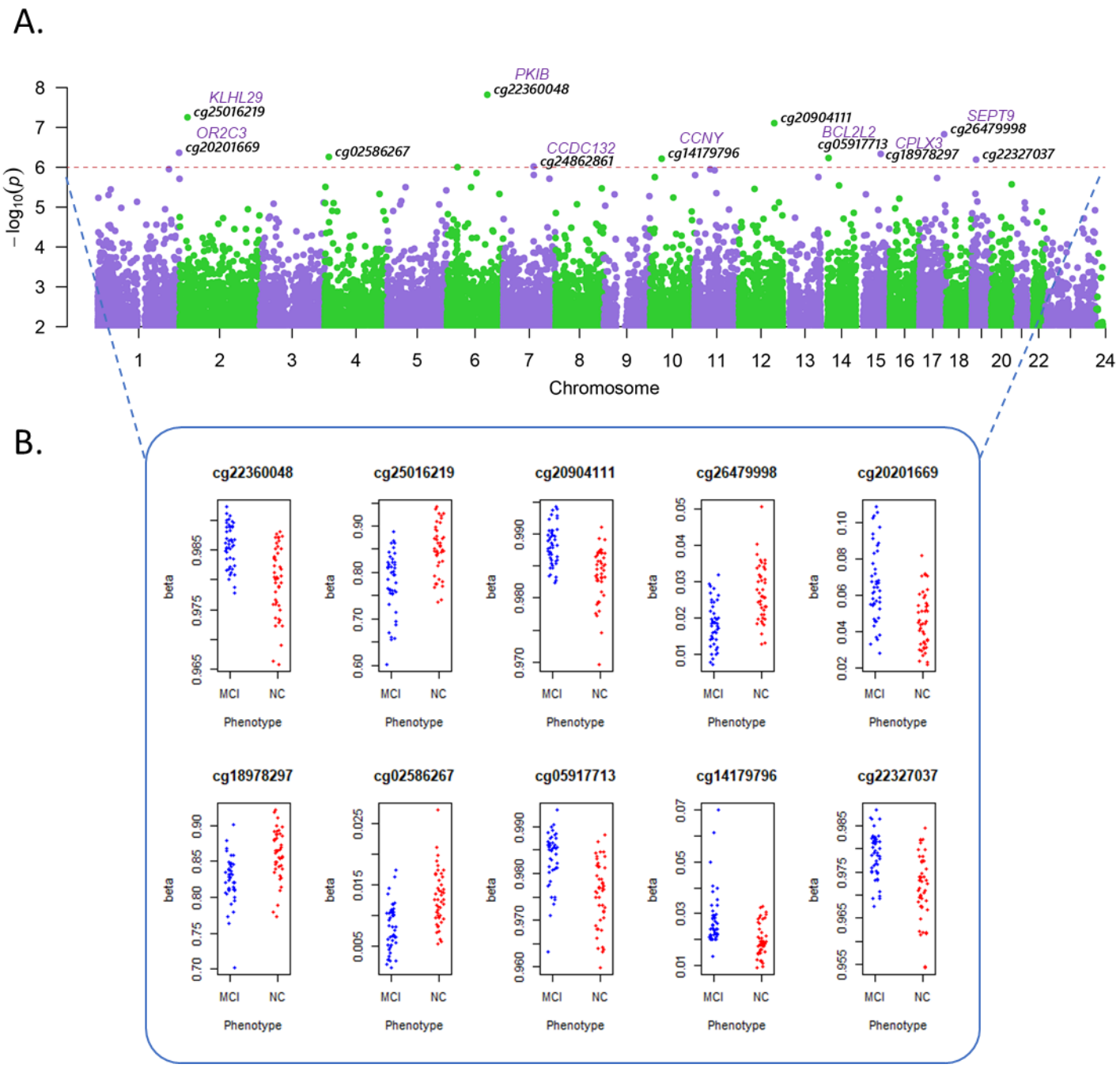


Figure1. ***(A) Manhattan plot highlighting differentially methylated probes****. The annotated genes using UCSC RefSeq are shown above the CpG ids.* ***(B) Scatter plots****. The distribution of beta intensity for each of the FDR significant cpg sites is shown between the NC (normal control) and MCI (mild cognitive impairment).*

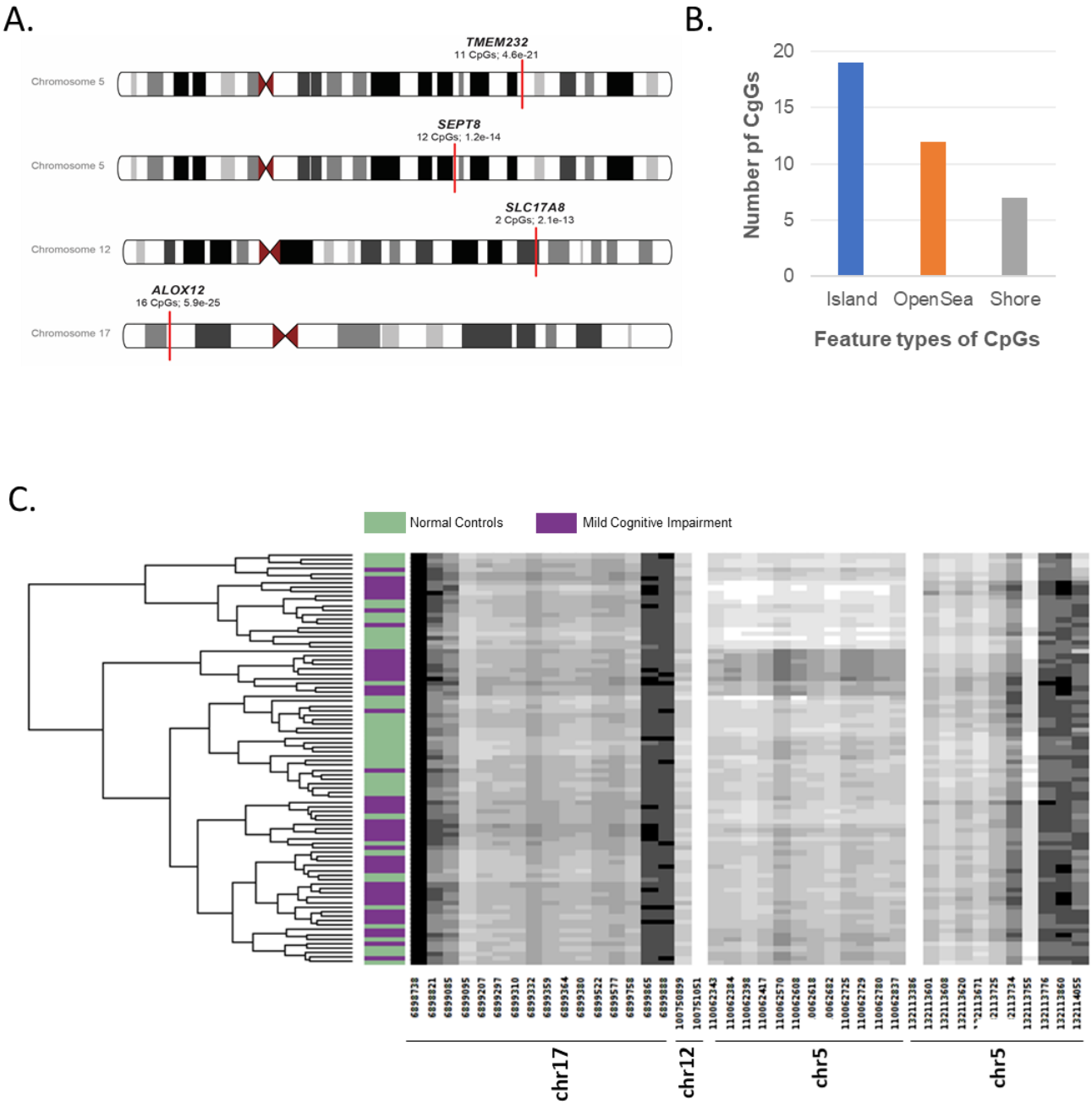


Figure 2. ***(A) Differentially Methylated Regions****. An ideogram representation of the location of differentially methylated regions and corresponding genes, number of CpGs in the regions and FDR p-value of the region.* ***(B) Classification of CpGs features****. An overview of all the CpGs sites in the identified regions based on their feature type. (C)* ***Dendrogram Plot****. Beta values for the probes within the differentially methylated regions for all samples*

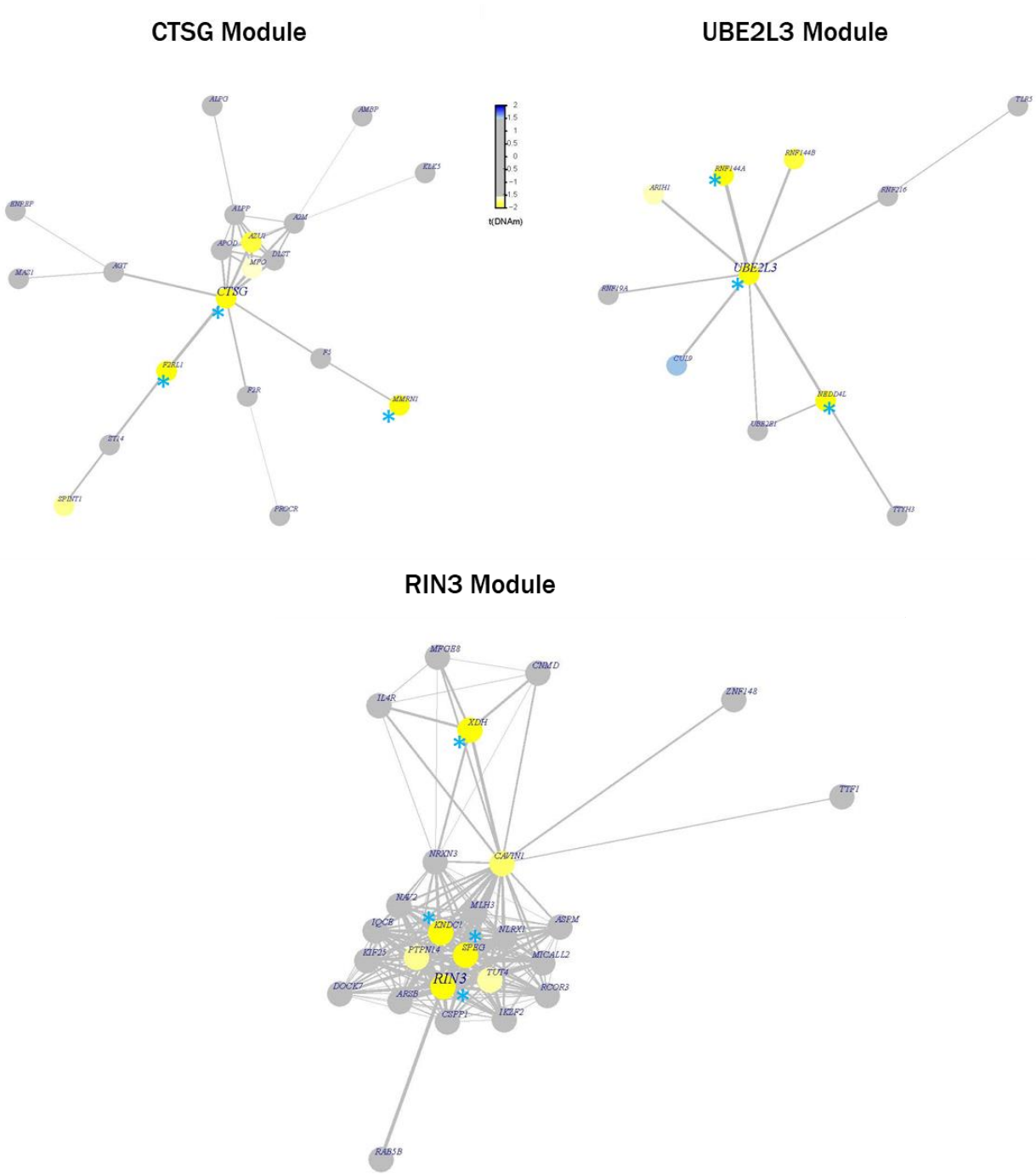


Figure 324. ***Differential Methylated Gene Modules.*** *The differentially methylated genes within the set are shown in yellow/blue on the legend; yellow signifies hypomethylated and blue signifies hypermethylated. Each of the module was found to be significant based on its overall p<0.05. Individual significance of each gene (p<0.05) in the module is highlighted with blue asterisk.*

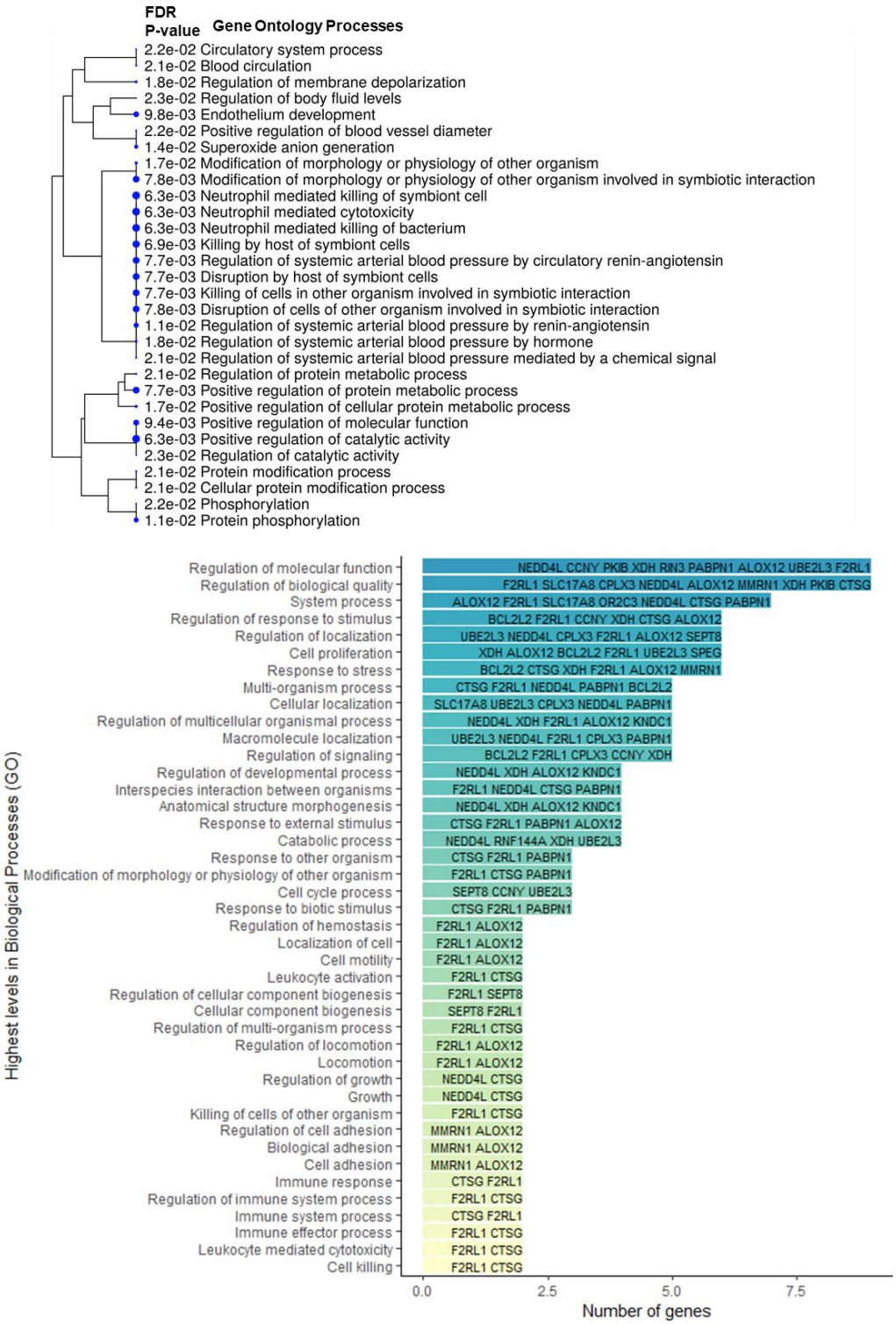


Figure 4. ***GO Enrichment*** *clustered together with their corresponding p-values. Bottom panel groups all identified genes in order of their highest level of process. The y-axis shows names of the processes, and x-axis shows number of genes, and each observation is labelled with the number of genes present in the respective gene category.*

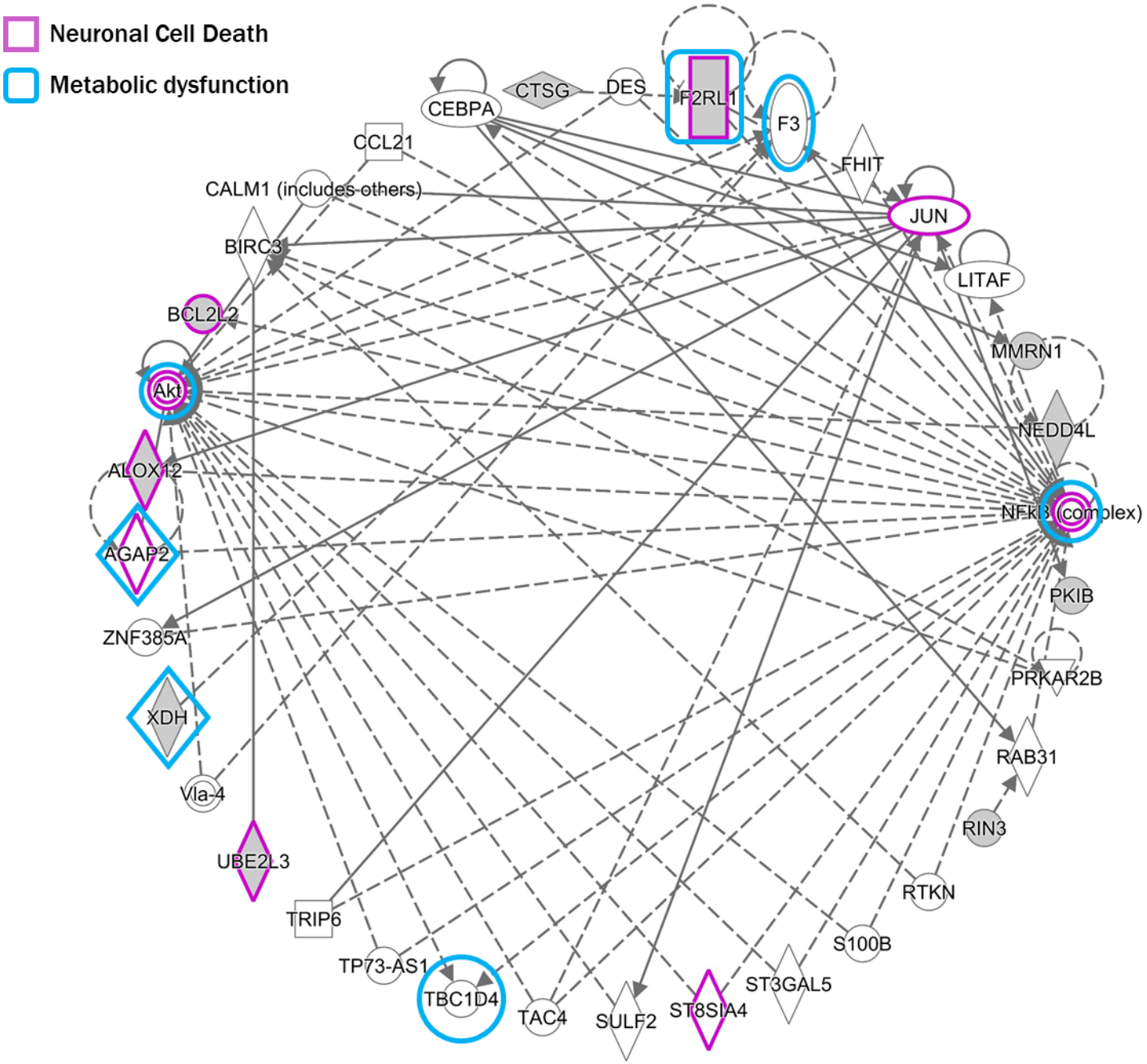


Figure 5. ***Pathway Enrichment****. Genes that were identified for neuronal cell death pathway are highlighted in pink and genes identified for metabolic dysfunction/insulin resistance are highlighted in blue.*

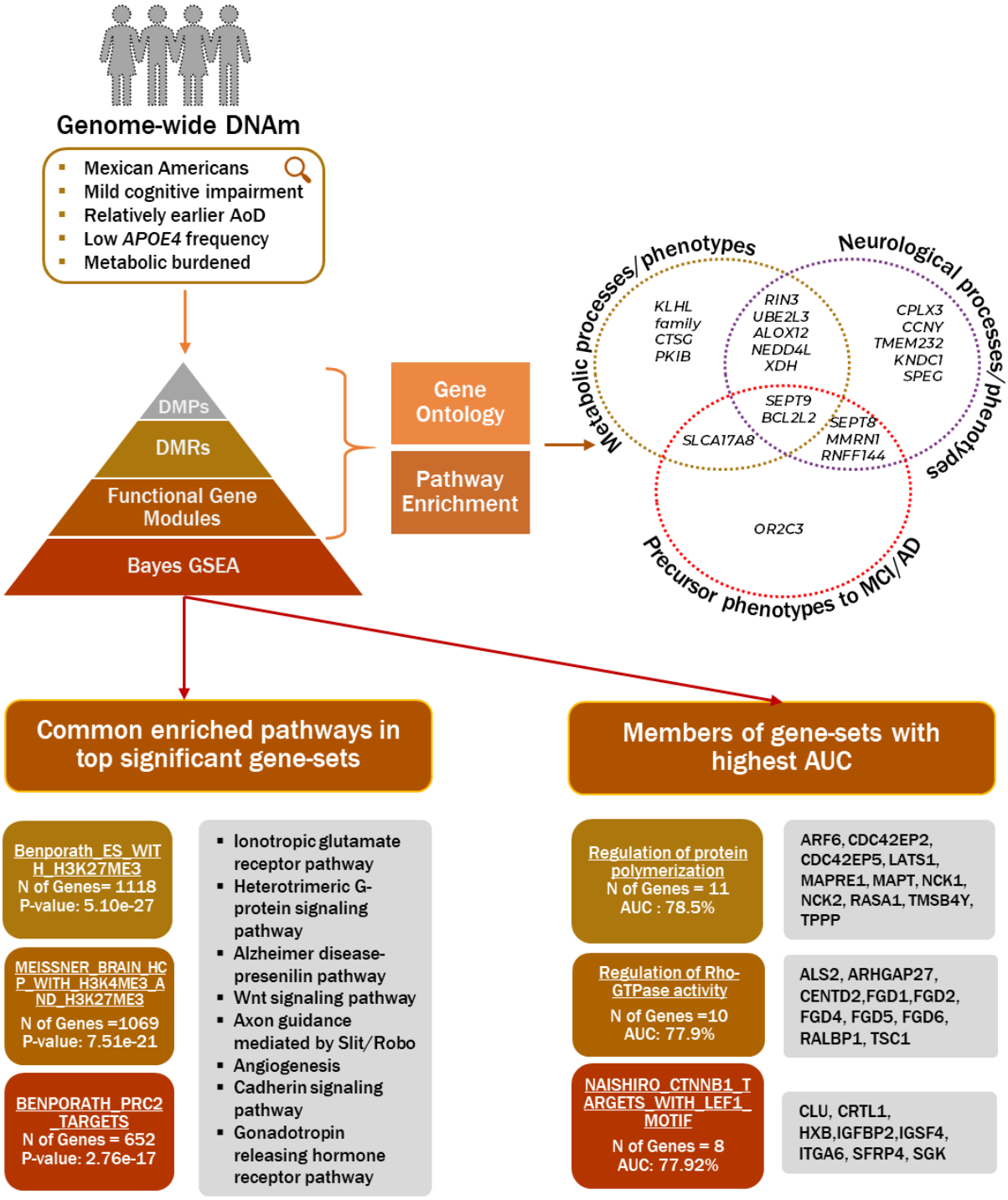


Figure 6. ***Visual Summary of identified genes and pathways****. Visual representation of methodology and an overview of findings converging on phenotypes and pathways.*

LEGEND TO FIGURES, TABLES AND SUPPLEMENTARY FILES

Tables

- **Table 1.** Demographic characteristics for the sample cohort

Figures

- **Figure 1.** (A) Manhattan plot highlighting differentially methylated probes. The annotated genes using UCSC RefSeq are shown above the CpG ids. (B) Scatter plots. The distribution of beta intensity for each of the FDR significant cpg sites is shown between the NC (normal control) and MCI (mild cognitive impairment).
- **Figure 2.** (A) Differentially Methylated Regions. An ideogram representation of the location of differentially methylated regions and corresponding genes, number of CpGs in the regions and FDR p-value of the region. (B) Classification of CpGs features. An overview of all the CpGs sites in the identified regions based on their feature type. (C) Dendrogram Plot. Beta values for the probes within the differentially methylated regions for all samples.
- **Figure 3.** Differential Methylated Gene Modules. The differentially methylated genes within the set are shown in yellow/blue on the legend; yellow signifies hypomethylated and blue signifies hypermethylated. Each of the module was found to be significant based on its overall $p<0.05$. Individual significance of each gene ($p<0.05$) in the module is highlighted with blue asterisk.
- **Figure 4.** GO Enrichment results for Biological Processes. The top panel shows the dendrogram of functionally similar GO processes clustered together with their corresponding p-values. Bottom panel groups all identified genes in order of their highest level of process. The y-axis shows names of the processes, and x-axis shows number of genes, and each observation is labelled with the number of genes present in the respective gene category.
- **Figure 5.** Pathway Enrichment. Genes that were identified for neuronal cell death pathway are highlighted in pink and genes identified for metabolic dysfunction/insulin resistance are highlighted in blue.
- **Figure 6**. Visual Summary of identified genes and pathways. Visual representation of methodology and an overview of findings converging on phenotypes, and pathways.

Supplementary Files

- Supplementary Figure 1
    - K (x-axis) is the standard error of the cross-validation error (y-axis) estimate. The lowest error is for K=1, indicating there is 1 population in the study. The values were calculated using SNPs in ADMIXTURE software (see methods for details)

- Supplementary Figure 2
    - The figure shows the number of publications from 1999-2019 for the PubMed query (last searched on June 18, 2019) mentioned in the subtitle. The y-axis shows shortened journal name and x-axis shows the year, the histogram overlaid on the axis shows the frequency of publications for that corresponding year.

- Supplementary file 1
    - **Table 1**. Differentially Methylated Probes between normal controls and mild cognitive impairment.
- Supplementary file 2
    - **Table 1**. Differentially Methylated Regions. The table shows the genomic coordinates of the region, CpG sites within the region and FDR p-value for the region.
    - **Table 2**. Details of individual probes reported within the differentially methylated regions as reported by DMRcate package.
    - **Table 3**. Annotation of the probes identified within differentially methylated regions.
- Supplementary file 3
    - **Table 1**. Differentially methylated modules/gene-sets and their corresponding p-values (P), number of genes (Size), and name of the genes in the respective sets (Genes).
    - **Table 2**. Differentially methylated genes in *RIN3* module/gene-set. List of all genes in the gene-set with their methylation values (stat(DNAm)) and corresponding p-values (P(DNAm)). Highlighted observations are significant genes based on $p<0.05$.
    - **Table 3**. Differentially methylated genes in *SPEG* module/gene-set. List of all genes in the gene-set with their methylation values (stat(DNAm)) and corresponding p-values (P(DNAm)). Highlighted observations are significant genes based on $p<0.05$.
    - **Table 4**. Differentially methylated genes in *CTSG* module/gene-set. List of all genes in the gene-set with their methylation values (stat(DNAm)) and corresponding p-values (P(DNAm)). Highlighted observations are significant genes based on $p<0.05$.
    - **Table 5**. Differentially methylated genes in *UBE2L3* module/gene-set. List of all genes in the gene-set with their methylation values (stat(DNAm)) and corresponding p-values (P(DNAm)). Highlighted observations are significant genes based on $p<0.05$.
- Supplementary file 4
    - **Table1**. Gene list used as input for gene set enrichment for gene ontology in ShinyGO tool.
    - **Table 2.** Result of enriched gene ontology biological processes.
    - **Table 3**. Result of input genes grouped by highest level of GO categories. Name of the process (High level GO category); number of genes in the category (N) and names of the gene members in the category (Genes).
    - **Table 4**. Details of enriched KEGG pathways.

- **Table 5**. Details of enriched pathways identified using IPA's (Ingenuity Pathway Analysis) knowledge base

- Supplementary file 5
  - **Table 1**. GSEA using empirical bayes. The table is ranked by P-value. Highlighted gene set names were explored for enriched pathways to identify biological meaning of identified gene sets. (see text for details; Figure 6)
  - **Table 2**. GSEA using empirical bayes. The table is ranked by AUC. Highlighted gene set names were explored for gene members from the MSigDB database and reported in Figure 6.
  - **Table 3**. The table shows the gene members of 'BENPORATH_ES_WITH_H3K27ME3', and enriched pathways from Panther's database (www.pantherdb.org) based on the genes in the set.
  - **Table 4.** The table shows the gene members of 'MEISSNER_BRAIN_HCP_WITH_H3K4ME3', and enriched pathways from Panther's database (www.pantherdb.org) based on the genes in the set.
  - **Table 5**. The table shows the gene members of 'BENPORATH_PRC2_TARGETS', and enriched pathways from Panther's database (www.pantherdb.org) based on the genes in the set.
  - **Table 6.** The table shows the gene members of 'REGULATION_OF_PROTEIN_POLYMERIZATION' from MSigDb.
  - **Table 7**. The table shows the gene members of 'REGULATION_OF_RHO_GTPASE_ACTIVITY' from MSigDb.
  - **Table 8**. The table shows the gene members of 'SHARMA_PILOCYTIC_ASTROCYTOMA_LOCATION_DN' from MSigDb.

ABBREVIATIONS

AD

ALZHEIMER'S DISEASE

MCI

MILD COGNITIVE IMPAIRMENT

NC

NORMAL CONTROLS

DMP

DIFFERENTIALLY METHYLATED PROBES

DMR

DIFFERENTIALLY METHYLATED REGIONS

IPA

INGENUITY PATHWAY ANALYSIS®

AUC

AREA UNDER THE CURVE

SNP

SINGLE NUCLEOTIDE POLYMORPHISMS

GSEA

GENE SET ENRICHMENT ANALYSIS

HABLE Health & Aging Brains of Latino Elders

MMSE

Mini Mental State Examination

IDF International Diabetes Federation

References

1

(2019) 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, **15**, 321-387.
2

Eshkoor, S.A., Hamid, T.A., Mun, C.Y. and Ng, C.K. (2015) Mild cognitive impairment and its management in older people. *Clin Interv Aging*, **10**, 687-693.
3

Gonyea, J.G., López, L.M. and Velásquez, E.H. (2016) The Effectiveness of a Culturally Sensitive Cognitive Behavioral Group Intervention for Latino Alzheimer's Caregivers. *Gerontologist*, **56**, 292-302.
4

O'Bryant, S.E., Johnson, L., Balldin, V., Edwards, M., Barber, R., Williams, B., Devous, M., Cushings, B., Knebl, J. and Hall, J. (2013) Characterization of Mexican Americans with mild cognitive impairment and Alzheimer's disease. *J Alzheimers Dis*, **33**, 373-379.
5

Johnson, L.A., Gamboa, A., Vintimilla, R., Cheatwood, A.J., Grant, A., Trivedi, A., Edwards, M., Hall, J.R. and O'Bryant, S.E. (2015) Comorbid Depression and Diabetes as a Risk for Mild Cognitive Impairment and Alzheimer's Disease in Elderly Mexican Americans. *J Alzheimers Dis*, **47**, 129-136.
6

Johnson, S.C., Dong, X., Vijg, J. and Suh, Y. (2015) Genetic evidence for common pathways in human age-related diseases. *Aging Cell*, **14**, 809-817.
7

Johnson, L.A., Large, S.E., Izurieta Munoz, H., Hall, J.R. and O'Bryant, S.E. (2019) Vascular

Depression and Cognition in Mexican Americans. *Dementia and geriatric cognitive disorders*, **47**, 68-78.
8

Vega, I.E., Cabrera, L.Y., Wygant, C.M., Velez-Ortiz, D. and Counts, S.E. (2017) Alzheimer's Disease in the Latino Community: Intersection of Genetics and Social Determinants of Health. *Journal of Alzheimer's disease : JAD*, **58**, 979-992.
9

Edwards, M., Hall, J., Williams, B., Johnson, L. and O'Bryant, S. (2016) Molecular markers of amnestic mild cognitive impairment among Mexican Americans. *J Alzheimers Dis*, **49**, 221-228.
10

O'Bryant, S.E., Xiao, G., Edwards, M., Devous, M., Gupta, V.B., Martins, R., Zhang, F. and Barber, R. (2013) Biomarkers of Alzheimer's disease among Mexican Americans. *J Alzheimers Dis*, **34**, 841-849.
11

Horvath, S. (2013) DNA methylation age of human tissues and cell types. *Genome biology*, **14**, R115.
12

Qazi, T.J., Quan, Z., Mir, A. and Qing, H. (2018) Epigenetics in Alzheimer's Disease: Perspective of DNA Methylation. *Mol Neurobiol*, **55**, 1026-1044.
13

Wang, S.C., Oelze, B. and Schumacher, A. (2008) Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS One*, **3**, e2698.
14

De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C. *et al.* (2014) Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nature neuroscience*, **17**, 1156-1163.
15

Pugazhenthi, S., Qin, L. and Reddy, P.H. (2017) Common neurodegenerative pathways in obesity, diabetes, and Alzheimer's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, **1863**, 1037-1045.
16

Wang, J., Gong, B., Zhao, W., Tang, C., Varghese, M., Nguyen, T., Bi, W., Bilski, A., Begum, S., Vempati, P. *et al.* (2014) Epigenetic Mechanisms Linking Diabetes and Synaptic Impairments. **63**, 645-654.
17

Braun, P.R., Han, S., Hing, B., Nagahama, Y., Gaul, L.N., Heinzman, J.T., Grossbach, A.J., Close,

L., Dlouhy, B.J., Howard, M.A. *et al.* (2019) Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Translational Psychiatry*, **9**, 47.
18

Marioni, R.E., McRae, A.F., Bressler, J., Colicino, E., Hannon, E., Li, S., Prada, D., Smith, J.A., Trevisi, L., Tsai, P.-C. *et al.* (2018) Meta-analysis of epigenome-wide association studies of cognitive abilities. *Molecular Psychiatry*, **23**, 2133-2144.
19

Teschendorff, A.E. and Relton, C.L. (2018) Statistical and integrative system-level analysis of DNA methylation data. *Nat Rev Genet*, **19**, 129-147.
20

Yates Kathy, F., Sweat, V., Yau Po, L., Turchiano Michael, M. and Convit, A. (2012) Impact of Metabolic Syndrome on Cognition and Brain. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **32**, 2060-2067.
21

Nilsson, E., Jansson, P.A., Perfilyev, A., Volkov, P., Pedersen, M., Svensson, M.K., Poulsen, P., Ribel-Madsen, R., Pedersen, N.L., Almgren, P. *et al.* (2014) Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes*, **63**, 2962-2976.
22

Lettre, G., Palmer, C.D., Young, T., Ejebe, K.G., Allayee, H., Benjamin, E.J., Bennett, F., Bowden, D.W., Chakravarti, A., Dreisbach, A. *et al.* (2011) Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet*, **7**, e1001300.
23

Hannibal, M.C., Ruzzo, E.K., Miller, L.R., Betz, B., Buchan, J.G., Knutzen, D.M., Barnett, K., Landsverk, M.L., Brice, A., LeGuern, E. *et al.* (2009) SEPT9 gene sequencing analysis reveals recurrent mutations in hereditary neuralgic amyotrophy. *Neurology*, **72**, 1755-1759.
24

Wain, L.V., Vaez, A., Jansen, R., Joehanes, R., van der Most, P.J., Erzurumluoglu, A.M., O'Reilly, P.F., Cabrera, C.P., Warren, H.R., Rose, L.M. *et al.* (2017) Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension*, in press.
25

Dayeh, T., Volkov, P., Salo, S., Hall, E., Nilsson, E., Olsson, A.H., Kirkpatrick, C.L., Wollheim, C.B., Eliasson, L., Ronn, T. *et al.* (2014) Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet*, **10**, e1004160.

26

Vaithianathan, T., Henry, D., Akmentin, W. and Matthews, G. (2015) Functional roles of complexin in neurotransmitter release at ribbon synapses of mouse retinal bipolar neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **35**, 4065-4070.
27

Annese, A., Manzari, C., Lionetti, C., Picardi, E., Horner, D.S., Chiara, M., Caratozzolo, M.F., Tullo, A., Fosso, B., Pesole, G. *et al.* (2018) Whole transcriptome profiling of Late-Onset Alzheimer's Disease patients provides insights into the molecular changes involved in the disease. *Scientific reports*, **8**, 4282-4282.
28

Haihua, Y. and Linghai, Y. (2016) Targeting cAMP/PKA pathway for glycemic control and type 2 diabetes therapy. *Journal of Molecular Endocrinology*, **57**, R93-R108.
29

Dagda, R.K. and Das Banerjee, T. (2015) Role of protein kinase A in regulating mitochondrial function and neuronal development: implications to neurodegenerative diseases. *Reviews in the neurosciences*, **26**, 359-370.
30

Banzhaf-Strathmann, J., Benito, E., May, S., Arzberger, T., Tahirovic, S., Kretzschmar, H., Fischer, A. and Edbauer, D. (2014) MicroRNA-125b induces tau hyperphosphorylation and cognitive deficits in Alzheimer's disease. *The EMBO journal*, **33**, 1667-1680.
31

Lee, M., Kwon, D.Y., Kim, M.S., Choi, C.R., Park, M.Y. and Kim, A.J. (2016) Genome-wide association study for the interaction between BMR and BMI in obese Korean women including overweight. *Nutrition research and practice*, **10**, 115-124.
32

Tavanez, J.P., Calado, P., Braga, J., Lafarga, M. and Carmo-Fonseca, M. (2005) In vivo aggregation properties of the nuclear poly(A)-binding protein PABPN1. *RNA*, **11**, 752-762.
33

Vasavada, M.M., Wang, J., Eslinger, P.J., Gill, D.J., Sun, X., Karunanayaka, P. and Yang, Q.X. (2015) Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment. *J Alzheimers Dis*, **45**, 947-958.
34

Ranzani, M., Iyer, V., Ibarra-Soria, X., Del Castillo Velasco-Herrera, M., Garnett, M., Logan, D. and Adams, D.J. (2017) Revisiting olfactory receptors as putative drivers of cancer. *Wellcome open research*, **2**, 9.
35

Zeng, L., Cai, C., Li, S., Wang, W., Li, Y., Chen, J., Zhu, X. and Zeng, Y.A. (2016) Essential Roles of Cyclin Y-Like 1 and Cyclin Y in Dividing Wnt-Responsive Mammary Stem/Progenitor Cells. *PLoS genetics*, **12**, e1006055-e1006055.
36

Park, M., Watanabe, S., Poon, Vivian Yi N., Ou, C.-Y., Jorgensen, Erik M. and Shen, K. (2011) CYY-1/Cyclin Y and CDK-5 Differentially Regulate Synapse Elimination and Formation for Rewiring Neural Circuits. *Neuron*, **70**, 742-757.
37

Folke, J., Pakkenberg, B. and Brudek, T. (2019) Impaired Wnt Signaling in the Prefrontal Cortex of Alzheimer's Disease. *Mol Neurobiol*, **56**, 873-891.
38

Conteh, A.M., Reissaus, C.A., Hernandez-Perez, M., Nakshatri, S., Anderson, R.M., Mirmira, R.G., Tersey, S.A. and Linnemann, A.K. (2019) Platelet-type 12-lipoxygenase deletion provokes a compensatory 12/15-lipoxygenase increase that exacerbates oxidative stress in mouse islet beta cells. *The Journal of biological chemistry*, **294**, 6612-6620.
39

Leibowitz, G., Kaiser, N. and Cerasi, E. (2011) β-Cell failure in type 2 diabetes. **2**, 82-91.
40

Khanna, S., Roy, S., Ryu, H., Bahadduri, P., Swaan, P.W., Ratan, R.R. and Sen, C.K. (2003) Molecular basis of vitamin E action: tocotrienol modulates 12-lipoxygenase, a key mediator of glutamate-induced neurodegeneration. *The Journal of biological chemistry*, **278**, 43508-43515.
41

Miller, M.W., Wolf, E.J., Sadeh, N., Logue, M., Spielberg, J.M., Hayes, J.P., Sperbeck, E., Schichman, S.A., Stone, A., Carter, W.C. *et al.* (2015) A novel locus in the oxidative stress-related gene ALOX12 moderates the association between PTSD and thickness of the prefrontal cortex. *Psychoneuroendocrinology*, **62**, 359-365.
42

Manohar, S., Dahar, K., Adler, H.J., Dalian, D. and Salvi, R. (2016) Noise-induced hearing loss: Neuropathic pain via Ntrk1 signaling. *Molecular and cellular neurosciences*, **75**, 101-112.
43

Haggstrom, J., Rosenhall, U., Hederstierna, C., Ostberg, P. and Idrizbegovic, E. (2018) A Longitudinal Study of Peripheral and Central Auditory Function in Alzheimer's Disease and in Mild Cognitive Impairment. *Dementia and geriatric cognitive disorders extra*, **8**, 393-401.
44

Marttinen, M., Kurkinen, K.M., Soininen, H., Haapasalo, A. and Hiltunen, M. (2015) Synaptic dysfunction and septin protein family members in neurodegenerative diseases. *Molecular neurodegeneration*, **10**, 16-16.

45

Yang, Y., Kim, J., Kim, H.Y., Ryoo, N., Lee, S., Kim, Y., Rhim, H. and Shin, Y.-K. (2015) Amyloid-β Oligomers May Impair SNARE-Mediated Exocytosis by Direct Binding to Syntaxin 1a. *Cell reports*, **12**, 1244-1251.
46

Rappaport, N., Nativ, N., Stelzer, G., Twik, M., Guan-Golan, Y., Stein, T.I., Bahir, I., Belinky, F., Morrey, C.P., Safran, M. *et al.* (2013) MalaCards: an integrated compendium for diseases and their annotation. *Database : the journal of biological databases and curation*, **2013**, bat018-bat018.
47

Chen, X., Long, F., Cai, B., Chen, X. and Chen, G. (2017) A novel relationship for schizophrenia, bipolar and major depressive disorder Part 5: a hint from chromosome 5 high density association screen. *American journal of translational research*, **9**, 2473-2491.
48

Boden, K.A., Barber, I.S., Clement, N., Patel, T., Guetta-Baranes, T., Brookes, K.J., Chappell, S., Craigon, J., Chapman, N.H., Consortium, A. *et al.* (2017) Methylation Profiling RIN3 and MEF2C Identifies Epigenetic Marks Associated with Sporadic Early Onset Alzheimer's Disease. *Journal of Alzheimer's disease reports*, **1**, 97-108.
49

Xu, W., Fang, F., Ding, J. and Wu, C. (2018) Dysregulation of Rab5-mediated endocytic pathways in Alzheimer's disease. *Traffic*, **19**, 253-262.
50

Stage, E., Duran, T., Risacher, S.L., Goukasian, N., Do, T.M., West, J.D., Wilhalme, H., Nho, K., Phillips, M., Elashoff, D. *et al.* (2016) The effect of the top 20 Alzheimer disease risk genes on gray-matter density and FDG PET brain metabolism. *Alzheimer's & dementia (Amsterdam, Netherlands)*, **5**, 53-66.
51

Yamada, Y., Sakuma, J., Takeuchi, I., Yasukochi, Y., Kato, K., Oguri, M., Fujimaki, T., Horibe, H., Muramatsu, M., Sawabe, M. *et al.* (2017) Identification of rs7350481 at chromosome 11q23.3 as a novel susceptibility locus for metabolic syndrome in Japanese individuals by an exome-wide association study. *Oncotarget*, **8**, 39296-39308.
52

Ji, J., Hao, Z., Liu, H., Liu, Y., Liu, J., Lin, B., Ma, C. and Lin, Y. (2018) Effect of KNDC1 overexpression on the senescence of human umbilical vein endothelial cells. *Molecular medicine reports*, **17**, 7037-7044.
53

Bernstein, A.I., Lin, Y., Street, R.C., Lin, L., Dai, Q., Yu, L., Bao, H., Gearing, M., Lah, J.J.,

Nelson, P.T. *et al.* (2016) 5-Hydroxymethylation-associated epigenetic modifiers of Alzheimer's disease modulate Tau-induced neurotoxicity. *Human molecular genetics*, **25**, 2437-2450.
54

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R. *et al.* (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, **50**, 1112-1121.
55

Kim, G.H., Kim, J.E., Rhie, S.J. and Yoon, S. (2015) The Role of Oxidative Stress in Neurodegenerative Diseases. *Exp Neurobiol*, **24**, 325-340.
56

Ahmad, W., Ijaz, B., Shabbiri, K., Ahmed, F. and Rehman, S. (2017) Oxidative toxicity in diabetes and Alzheimer's disease: mechanisms behind ROS/ RNS generation. *Journal of biomedical science*, **24**, 76.
57

Aliciguzel, Y., Ozen, I., Aslan, M. and Karayalcin, U. (2003) Activities of xanthine oxidoreductase and antioxidant enzymes in different tissues of diabetic rats. *The Journal of laboratory and clinical medicine*, **142**, 172-177.
58

Stock, A.J., Kasus-Jacobi, A. and Pereira, H.A. (2018) The role of neutrophil granule proteins in neuroinflammation and Alzheimer's disease. *J Neuroinflammation*, **15**, 240-240.
59

Csiszar, A., Tucsek, Z., Toth, P., Sosnowska, D., Gautam, T., Koller, A., Deak, F., Sonntag, W.E. and Ungvari, Z. (2013) Synergistic effects of hypertension and aging on cognitive function and hippocampal expression of genes involved in β-amyloid generation and Alzheimer's disease. *Am J Physiol Heart Circ Physiol*, **305**, H1120-H1130.
60

Kempuraj, D., Thangavel, R., Natteru, P.A., Selvakumar, G.P., Saeed, D., Zahoor, H., Zaheer, S., Iyer, S.S. and Zaheer, A. (2016) Neuroinflammation Induces Neurodegeneration. *Journal of neurology, neurosurgery and spine*, **1**.
61

Kempuraj, D., Selvakumar, G.P., Thangavel, R., Ahmed, M.E., Zaheer, S., Kumar, K.K., Yelam, A., Kaur, H., Dubova, I., Raikwar, S.P. *et al.* (2018) Glia Maturation Factor and Mast Cell-Dependent Expression of Inflammatory Mediators and Proteinase Activated Receptor-2 in Neuroinflammation. *J Alzheimers Dis*, **66**, 1117-1129.
62

Lim, J., Iyer, A., Liu, L., Suen, J.Y., Lohman, R.J., Seow, V., Yau, M.K., Brown, L. and Fairlie,

D.P. (2013) Diet-induced obesity, adipose inflammation, and metabolic dysfunction correlating with PAR2 expression are attenuated by PAR2 antagonism. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, **27**, 4757-4767.
63

Nishioka, K., Hayashi, S., Farrer, M.J., Singleton, A.B., Yoshino, H., Imai, H., Kitami, T., Sato, K., Kuroda, R., Tomiyama, H. *et al.* (2006) Clinical heterogeneity of α-synuclein gene duplication in Parkinson's disease. **59**, 298-309.
64

Zyskind, J.W., Wang, Y., Cho, G., Ting, J.H., Kolson, D.L., Lynch, D.R. and Jordan-Sciutto, K.L. (2015) E2F1 in neurons is cleaved by calpain in an NMDA receptor-dependent manner in a model of HIV-induced neurotoxicity. *Journal of neurochemistry*, **132**, 742-755.
65

Denechaud, P.D., Fajas, L. and Giralt, A. (2017) E2F1, a Novel Regulator of Metabolism. *Frontiers in endocrinology*, **8**, 311.
66

Russo, C.J., Melista, E., Cui, J., DeStefano, A.L., Bakris, G.L., Manolis, A.J., Gavras, H. and Baldwin, C.T. (2005) Association of NEDD4L ubiquitin ligase with essential hypertension. *Hypertension*, **46**, 488-491.
67

Nilsson, E., Matte, A., Perfilyev, A., de Mello, V.D., Käkelä, P., Pihlajamäki, J. and Ling, C. (2015) Epigenetic Alterations in Human Liver From Subjects With Type 2 Diabetes in Parallel With Reduced Folate Levels. *J Clin Endocrinol Metab*, **100**, E1491-E1501.
68

Atkin, G. and Paulson, H. (2014) Ubiquitin pathways in neurodegenerative disease. **7**.
69

Mellendijk, L., Wiesmann, M. and Kiliaan, A.J. (2015) Impact of Nutrition on Cerebral Circulation and Cognition in the Metabolic Syndrome. *Nutrients*, **7**, 9416-9439.
70

Li, M., Quan, C., Toth, R., Campbell, D.G., MacKintosh, C., Wang, H.Y. and Chen, S. (2015) Fasting and Systemic Insulin Signaling Regulate Phosphorylation of Brain Proteins That Modulate Cell Morphology and Link to Neurological Disorders. *The Journal of biological chemistry*, **290**, 30030-30041.
71

Massaad, C.A. and Klann, E. (2011) Reactive oxygen species in the regulation of synaptic plasticity and memory. *Antioxid Redox Signal*, **14**, 2013-2054.
72

Ferreira, L.S.S., Fernandes, C.S., Vieira, M.N.N. and De Felice, F.G. (2018) Insulin Resistance in Alzheimer's Disease. *Front Neurosci*, **12**, 830-830.
73

Pruimboom, L., Raison, C.L. and Muskiet, F.A.J. (2015) Physical Activity Protects the Human Brain against Metabolic Stress Induced by a Postprandial and Chronic Inflammation %J Behavioural Neurology. **2015**, 11.
74

Mottahedin, A., Ardalan, M., Chumak, T., Riebe, I., Ek, J. and Mallard, C. (2017) Effect of Neuroinflammation on Synaptic Organization and Function in the Developing Brain: Implications for Neurodevelopmental and Neurodegenerative Disorders. *Front Cell Neurosci*, **11**, 190-190.
75

Li, P., Marshall, L., Oh, G., Jakubowski, J.L., Groot, D., He, Y., Wang, T., Petronis, A. and Labrie, V. (2019) Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms. *Nature Communications*, **10**, 2246.
76

Folstein, M.F., Folstein, S.E. and McHugh, P.R. (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189-198.
77

Morris, J.C. (1993) The Clinical Dementia Rating (CDR): current version and scoring rules.[see comment]. *Neurology.*, **43**, 2412-2414.
78

O'Bryant, S.E., Lacritz, L.H., Hall, J., Waring, S.C., Chan, W., Khodr, Z.G., Massman, P.J., Hobson, V. and Cullum, C.M. (2010) Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. *Arch Neurol*, **67**, 746-749.
79

O'Bryant, S.E., Waring, S.C., Cullum, C.M., Hall, J., Lacritz, L., Massman, P.J., Lupo, P.J., Reisch, J.S. and Doody, R. (2008) Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study. *Arch Neurol*, **65**, 1091-1095.
80

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack Jr, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R. *et al.* (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, **7**, 263-269.

81

Petersen, R.C. (2003). Oxford University Press, New York, in press.
82

Ivnik, R., Malec, JF, Smith, GE, Tangalos, EG, Petersen, RC, Kokmen, E, & Kurkland, LT. (1992) Mayo's Older Americans Normative Studies: WAIS-R norms for age 56 to 97. *The Clinical Neuropsychologist*, **6**, 1-30.
83

Alberti, K.G., Zimmet, P. and Shaw, J. (2005) The metabolic syndrome--a new worldwide definition. *Lancet (London, England)*, **366**, 1059-1062.
84

Jackson, R.T., Sarafrazi, N. and Momen, B. (2011) Appropriate Waist Circumference Cutoff Values for the Diagnosis of Metabolic Syndrome in Mexican American Adults. **25**, lb243-lb243.
85

Masconi, K.L., Matsha, T.E., Erasmus, R.T. and Kengne, A.P. (2015) Effects of Different Missing Data Imputation Techniques on the Performance of Undiagnosed Diabetes Risk Prediction Models in a Mixed-Ancestry Population of South Africa. *PLOS ONE*, **10**, e0139210.
86

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.
87

Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, **19**, 1655-1664.
88

Yang, Z., Tian, Y., Teschendorff, A.E., Webster, A.P., Feber, A., Beck, S. and Morris, T.J. (2017) ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*, **33**, 3982-3984.
89

Ritchie, M.E., Smyth, G.K., Phipson, B., Wu, D., Hu, Y., Shi, W. and Law, C.W. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**, e47-e47.
90

Peters, T.J., Buckley, M.J., Statham, A.L., Pidsley, R., Samaras, K., R, V.L., Clark, S.J. and Molloy, P.L. (2015) De novo identification of differentially methylated regions in the human genome. *Epigenetics & chromatin*, **8**, 6.
91

Jiao, Y., Widschwendter, M. and Teschendorff, A.E. (2014) A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinformatics*, **30**, 2360-2366.
92

Marioni, R., Harris, S.E., McRae, A.F., Zhang, Q., Hagenaars, S.P., Hill, W.D., Davies, G., Ritchie, C.W., Gale, C., Starr, J.M. *et al.* (2018), in press.
93

Hahne, F. and Ivanek, R. (2016) Mathé, E. and Davis, S. (eds.), In *Statistical Genomics: Methods and Protocols*. Springer New York, New York, NY, in press., pp. 335-351.
94

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer Publishing Company, Incorporated.