SYSTEMATIC IDENTIFICATION OF RISK FACTORS FOR ALZHEIMER'S DISEASE THROUGH SHARED GENETIC ARCHITECTURE AND ELECTRONIC MEDICAL RECORDS

LI LI*
Div. of Systems Medicine, Dept. of Pediatrics, Stanford University School of Medicine
Stanford, CA 94305, USA
Email: l3li@stanford.edu

DAVID RUAU*
Div. of Systems Medicine, Dept. of Pediatrics, Stanford University School of Medicine
Stanford, CA 94305, USA
Email: druau@stanford.edu

RONG CHEN
Personalis Inc., 1350 Willow Rd.,
Menlo Park, CA 94025, USA
Email: rong.chen@personalis.com

SUSAN WEBER
Stanford Center for Clinical Informatics, Stanford University School of Medicine
Stanford, CA 94305, USA
Email: scweber@stanford.edu

ATUL J. BUTTE
Div. of Systems Medicine, Dept. of Pediatrics, Stanford University School of Medicine
Stanford, CA 94305, USA
Email: abutte@stanford.edu

**: authors are equally contributed to this work*

Alzheimer's disease (AD) is one of the leading causes of death for older people in US with rapidly increasing incidence. AD irreversibly and progressively damages the brain, but there are treatments in clinical trials to potentially slow the development of AD. We hypothesize that the presence of clinical traits, sharing common genetic variants with AD, could be used as a non-invasive means to predict AD or trigger for administration of preventative therapeutics. We developed a method to compare the genetic architecture between AD and traits from prior GWAS studies. Six clinical traits were significantly associated with AD, capturing 5 known risk factors and 1 novel association: erythrocyte sedimentation rate (ESR). The association of ESR with AD was then validated using Electronic Medical Records (EMR) collected from Stanford Hospital and Clinics. We found that female patients and with abnormally elevated ESR were significantly associated with higher risk of AD diagnosis (OR: 1.85 [1.32-2.61], p=0.003), within 1 year prior to AD diagnosis (OR: 2.31 [1.06-5.01], p=0.032), and within 1 year after AD diagnosis (OR: 3.49 [1.93-6.31], p<0.0001). Additionally, significantly higher ESR values persist for all time courses analyzed. Our results suggest that ESR should be tested in a specific longitudinal study for association with AD diagnosis, and if positive, could be used as a prognostic marker.

1. Introduction

Alzheimer's disease (AD) is the fifth-leading cause of death in older people and is the most common cause of dementia (up to 75%), with an approximately 26 million affected individuals worldwide estimated to reach 115 million by 2050 (*1-3*). Of those with Alzheimer's disease, an estimated 4% are under age 65, 6 % are 65 to 74, 44 % are 75 to 84, and 46% are 85 or older (*2*). Compared with men, women have a 1.54 fold increased risk for AD (95% CI, 1.21 to 1.96) (*4*).

About 25% of all AD cases have familial history (i.e., with 2 or more persons in a family having AD). Nevertheless, the main cause remains unknown, which may due to genetic and environment factors (*5*). AD is an irreversible and progressive brain disease, which can be diagnosed using behavioral observations and the gold standard for confirmation rely on neuropathologic findings of beta-amyloid plaques and intraneuronal neurofibrillary tangles upon autopsy examination (*6*). Therefore, identifying clinical manifestations of risk factors related with AD are critically needed for early diagnosis, prognostics and preventive care of AD. Currently, the known risk factors of AD are advancing age, family history, gender, *APOE* ε4 allelic variant, cardiovascular factors, mild cognitive impairment, life style, and head trauma, which were investigated through large scale epidemiological studies (*7-13*). However, these factors have relatively weak predictive effects. It is still necessary to find more potential risk factors which may contribute to AD development (*3*).

Over the past decade, Genome-Wide Association Study (GWAS) and candidate gene studies have identified genetic variants for thousands of diseases and traits (*14-16*). A previous study has shown the "human disease network" where two diseases were connected to each other if they shared at least one gene from Online Mendelian Inheritance in Man (OMIM), however, they did not integrate GWAS studies (*17*). We hypothesize that traits from GWAS studies might serve as additional risk factors for disease, here specifically looking at AD. We theorize that if a prior GWAS for a trait has identified a list of genes with variants that significantly match the list of genes with variants associated with AD, then that trait might serve as a predictive factor for AD.

In this study, we used those variants and develop a method to systemically identify associations between clinical traits and AD in a fast and efficient way. We searched for traits sharing common genetic variants with AD that could serve as a means to prognose AD, and possibly provide opportunities for life-style interventions and preventive drug treatment. We validated our novel finding using Electronic Medical Records (EMR) through an independent large patient cohort with more than 15,000 patients from Stanford Hospital and Clinics (SHC) (*18*).

2. Methods

2.1 Utilizing VARiant Informing MEDicine (VARIMED)

The overall experiment design is shown in Figure 1. GWAS have enabled the elucidation of the genetic architecture of hundreds of diseases, many of which are polygenic complex disorders. We have manually curated a unique database called VARiant Informing MEDicine (VARIMED) (*19*), holding manually curated, quantitative human disease-SNP associations extracted from the full text, figures, tables, and supplemental materials of human genetic related publications.

VARIMED is a comprehensive genetic association database with over 100 features stored including diseases (e.g. diabetes, lung cancer), clinical traits (e.g. blood pressure, creatinine levels), gene symbol, dbSNP, odds ratio, and published p-value of association from literature (*19-22*). Diseases are categorized and currently mapped to Concept Unique Identifiers (CUI) from the Unified Medical Language System. All the genetic variants (SNPs) were systematically annotated to the genes with the most recent NCBI Entrez gene identifiers using Entrez dbSNP by AILUN (*23*). At the time of this writing, VARIMED covers 8,962 human genetics papers from GWAS and candidate gene studies, including 87,553 SNPs annotated to 8,913 genes for 1,119 diseases and 1,257 clinical traits.

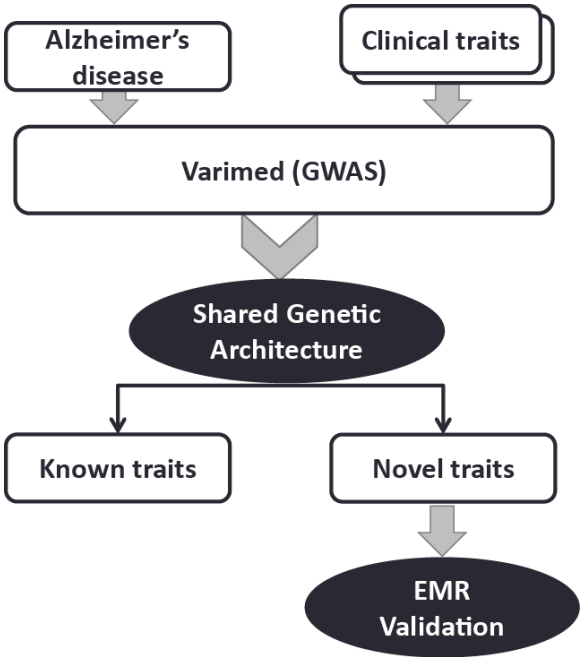


Figure 1: Work flow for entire experiment design

2.2 Assessing shared genetic architecture for Alzheimer's disease (AD) and clinical traits

We compared the shared genetic architecture for all available clinical traits in VARIMED with against Alzheimer's disease (AD) by first collecting all genetic variants related with AD and 1,257 traits. We selected only those variants associated at the gene level with AD and traits with $p \leq 1E\text{-}8$ as a highly stringent threshold to reduce the chance of false positive results.

As some genes could be shared solely between a few traits, and other genes shared across thousands, we needed an approach to capture the specificity and relevance of the genetic association. We used a Term Frequency–Inverse Document Frequency (TF-IDF) weighing method (*24*) to take into account the popularity of the genes. The detailed calculation procedure is as follows. First, we calculated a term frequency (TF) using:

$$tf(i,j) = \frac{frequency\ of\ the\ gene\ i\ in\ phenotype\ j}{number\ of\ all\ genes\ in\ phenotype\ j} \quad (1)$$

where phenotype refers to a trait or the disease AD.

The *tf* score indicates the occurrence frequency level of gene *i* in phenotype *j*, similar to a precision measure. Then, we calculated the inverse document frequency (IDF) using:

$$idf(i) = \log_{10}\left(\frac{total\ number\ of\ phenotypes}{number\ of\ phentoypes\ containing\ gene\ i}\right) \quad (2)$$

A larger *idf* score implies a lower popularity of gene *i* among the phenotypes (akin to a higher accuracy), which gives more weight to the gene as it might only be shared between these two phenotypes. Last, the TF-IDF score was calculated using:

$$tf\text{-}idf_{(i,j)} = tf_{(i,j)} \times idf_i \quad (3)$$

A high weight in *tf-idf* is reached by a high gene frequency (in the given phenotype) and a low phenotype frequency of the gene across all phenotypes studied.

Thus, for every AD-trait pairs a TF-IDF score for every shared gene was computed. The similarity between AD and all traits was then estimated by the cosine distance based on *tf-idf* scores.

To evaluate the statistical significance of the distance scores obtained, we computed the False Discovery Rate (FDR) by random shuffling (1,000 times) the genes across all the traits and re-computing the AD-trait distance. The q-value was calculated as the ratio of the expected number of false positive over the total number of hypothesis tested (*25*). Q-value ≤ 0.01 was selected as threshold of significant association between AD and trait pairs.

2.3 Validation of novel finding from the independent electronic medical records

To assess the clinical relevance of our novel finding, we used electronic medical records (EMR) data extracted from Stanford Translational Research Integrated Database Environment (STRIDE). STRIDE is a research and development project at Stanford University to create a standards-based informatics platform supporting clinical and translational research (*18*). STRIDE contains a clinical data warehouse which is comprised of comprehensive clinical information such as ICD9 diagnoses codes, CPT procedure codes, and lab results on over 1.7 million pediatric and adult patients cared for at Stanford Hospital and Clinic. STRIDE has been implemented at SHC since 2005. We used patient data in STRIDE as an independent cohort specifically recruited for this study to validate the hypothetical associations observed between AD and traits at the genetic level. Patients with AD were retrieved using the ICD9 code = 331.0, the rest of the hospital population being considered as control.

Chi-square test and Mann–Whitney U test were used to investigate the effect of the traits and AD. All statistics and graphs were carried out by SAS 9.2 (SAS institute Inc., Cary, SC) and R 2.15.0 (*26*).

2.4 Ethical statement

Data collected from STRIDE did not contain any protected health information and thus the study was considered non-human subjects' research, as determined and approved by the Institutional Review Board at Stanford.

3 Result

3.1 Discovering genetic architecture related with Alzheimer's disease

From 8,962 GWAS and candidate genes studies implemented in VARIMED, we queried the number of unique SNPs, genes, and genetic studies associated with Alzheimer's disease (AD). We used a stringent and well-accepted p-value threshold ≤ 1E-8 as genome wide significant, and identified 89 SNPs within 28 genes published across 44 genetic studies associated (Table 1).

Table 1: Genes and number of genetic studies associated with Alzheimer's disease

Gene	SNP Count	P-value	Study Count
APOD	1	0	1
SORCS1	1	0	2
APOC1	1	1.00E-300	8
TOMM40	9	1.28E-299	10
PVRL2	18	5.65E-74	7
APOE	2	1.83E-67	8
BCL3	2	1.93E-21	3
ABCA7	1	5.00E-21	3
LRRC68	4	2.16E-20	2
BCAM	1	5.54E-19	1
CLU	2	1.10E-16	6
MS4A6A	6	1.20E-16	2
PCK1	1	2.00E-16	4
ZNF224	1	2.00E-16	4
CR1	7	3.70E-14	5
PVR	1	6.17E-12	2
NKPD1	1	1.04E-11	1
MS4A4A	2	4.71E-11	1
GAB2	3	9.66E-11	5
MTHFD1L	1	1.90E-10	2
CALHM1	1	2.00E-10	3
CLPTM1	1	2.00E-10	1
CEACAM16	1	7.68E-10	2
PICALM	13	9.57E-10	1
CD33	1	1.60E-09	2
MS4A4E	5	1.98E-09	1
MS4A2	1	2.94E-09	2
CD2AP	1	8.60E-09	2

3.2 Systematically identifying the significant traits with genetic architecture shared with AD

We identified 249 traits where at least one gene was genetically associated. In our study, a trait was defined as a human-related physical or cognitive measurement, which was not explicitly a predisposition to another disease. To evaluate the significance of the shared variants in AD and all possible trait pairings, we attributed to each gene a measure based on their popularity using TF-IDF weight adjustment, and tested for significance using random permutation (see Methods

section 2.2). We identified 6 significant traits that paired with AD with q-value ≤ 0.01 (Table 2) based on the method we described above. All 6 traits originated from different published GWAS studies, suggesting that integrating different GWAS studies to discover underlying shared genetic architecture between diseases and traits can yield novel risk factors for the disease.

Among the 6 traits, 5 were related lipid tests and all shared variants in *APOC1*, *PVRL2*, and *TOMM40* genes in their genetics. *APOE* was shared in the lipid panel however was absent in Lipoprotein-Associated phospholipase a2 activity (Lp-PLA2) (Table 2). Erythrocyte sedimentation rate (ESR), a common immunology test to measure non-specific inflammation showed significant genetic association with AD through only one gene: complement component (3b/4b) receptor 1 (*CR1*). *CR1* was associated with ESR and AD solely and not with other phenotypes in VARIMED. *CR1* is a receptor and binds to *C3* and *C4* complement genes, which have been shown an increase in chronic inflammation (*27*), in risk of developing a myocardial infarction (*28*), and in deceased donor who progressed poor graft function due to cold ischemic injury with potential inflammation after kidney transplantation (*29*).

Among the 6 traits associated with AD, 5 associations were already known to be either risk factors or comorbidities of AD in the published literature (Table 2). Lipoprotein-Associated phospholipase a2 (Lp-PLA2) is a risk factor associated with the risk of dementia in the Rotterdam study, independently of cardiovascular and inflammatory factors (*30*). C-reactive protein (CRP) level is a risk factor where elevated CRP continues to predict increased dementia severity suggesting a possible proinflammatory endophenotype in AD (*31*). In addition, lipid level has been seen to increase in patients who have already developed AD. Apolipoprotein b (ApoB) level is increased in AD patients, suggesting that ApoE may not be the single factor in lipid metabolism to play a role in AD pathogenesis (*32*). Higher total cholesterol and LDL levels were significantly related to pathologically defined AD, which in turn suggests serum lipids have a role in the pathogenesis of AD and interventions may modify the progression of disease (*33*,*34*). Furthermore, the shared genes also explain the genetic cause between AD and these 5 traits.

Table 2: Clinical traits significant associated with Alzheimer's disease

Clinical Trait	Gene Count	Common Genes	Gene Shared	Q-value	Reference
Lipoprotein-Associated phospholipase a2 activity	12	3	*APOC1;PVRL2; TOMM40*	< 0.001	*30*
Apolipoprotein b levels	12	4	*APOC1; APOE; PVRL2; TOMM40*	< 0.001	*32*
C reactive protein levels	17	3	*APOC1; APOE; TOMM40*	0.002	*31*
LDL cholesterol levels	44	4	*APOC1; APOE; PVRL2; TOMM40*	0.002	*34*
Erythrocyte sedimentation rate	5	1	*CR1*	0.004	Novel
Cholesterol levels	50	4	*APOC1; APOE; PVRL2; TOMM40*	0.004	*33*

3.3 Clinical validation for novel trait ESR association with AD in an independent cohort

We identified ESR as a novel trait significantly sharing genes with genetic variants with AD. Since ESR is a well-known clinical measurement and non-specific marker of inflammation, and not known to be associated with AD, we evaluated the hypothesis that ESR might be abnormal

before the diagnosis of AD. We obtained all ESR lab results from Stanford Hospital and Clinics from 2005 until July 15, 2012 for patients with and without an AD diagnosis. Our case cohort was constituted of 212 patients who were ever measured for ESR and had at least one diagnosis code of AD (mean age 81±8; range [48-96]) with 135 females and 78 males. We considered patients older than age 50, having at least one measurement of ESR, and never having a diagnosis code of AD as the control group, resulting in 15,040 unique patients. Reference ranges for Erythrocyte sedimentation rate (ESR) lab tests were defined 0-20 mm/hr for female <50, 0-30 mm/hr for female ≥ 50 years, and 0-20 mm/hr for male ≥ 50 years based on MedlinePlus (http://www.nlm.nih.gov/medlineplus/).

As AD is known to exhibit a sex difference in prevalence (*4*), we evaluated each gender separately. First, we compare the abnormal high ESR percentage for AD and control patients across all available time points (ESR measurement irrespective of the AD diagnosis code(s)) to test the overall association. Then, we compared the abnormal high ESR percentage within 1 year prior to our first diagnosis code of AD in AD patients, and first diagnosis code of anything other than AD in control patients, to investigate whether changes in ESR could be a risk factor to predict the AD incidence. Finally, we compared the ESR within 1 year after our first diagnosis code of AD in AD patients, and first diagnosis of anything other than AD in control patients, to evaluate whether ESR changes could be a consequence of the AD diagnosis.

In female, patients with abnormally high ESR (45%) (> 30 mm/hr) were significantly associated with having a diagnosis code of AD irrespective of lab and diagnosis timing (OR: 1.85 [1.32-2.61], p=0.0003). The effect was strengthened when looking at ESR measurements within 1 year prior to our first AD diagnosis for patients (OR: 2.31 [1.06-5.01], p=0.032), and within 1 year after our first AD diagnosis on patients (Table 3). Furthermore, ESR values were significantly higher across all time points (p<0.0001), within 1 year prior to diagnosis (p=0.0025), and within 1 year after diagnosis (p<0.0001) in AD versus controls by Mann–Whitney U test (Figure 1A).

Table 3: Clinical validation through electronic medical record from STRIDE by Chi-square test

Time Frame	Gender	OR (95%CI)	% in each cohort having an abnormal high ESR (%, AD vs. Control)	P (Chi-square)	# of AD	# of Control	Total #
All time points, irrespective of diagnosis timing	F	1.85 (1.32-2.61)	53.33% vs. 38.15%	0.0003	135	8769	8904
	M	1.42 (0.91-2.23)	56.41% vs. 47.60%	0.1216	78	6271	6349
ESR testing 1 year prior our first diagnosis	F	2.31 (1.06-5.01)	44.74% vs. 25.96%	0.032	38	104	142
	M	2.41 (0.94-6.18)	54.17% vs. 32.88%	0.0625	24	73	97
ESR testing 1 year after our first diagnosis	F	3.49 (1.93-6.31)	69.23% vs. 39.20%	<.0001	52	3194	3246
	M	1.79 (0.83-3.84)	52.79% vs. 66.67%	0.1302	30	2487	2517

In males, patient with high ESR (54%) (> 20 mm/hr) show a trend towards association with AD within 1 year prior to our patients' first AD diagnosis code (OR: 2.41 [0.94-6.18], p=0.0625) (Table 3). ESR values were overall significantly higher compared with control (p=0.0198) by Mann–Whitney U test (Figure 1B).

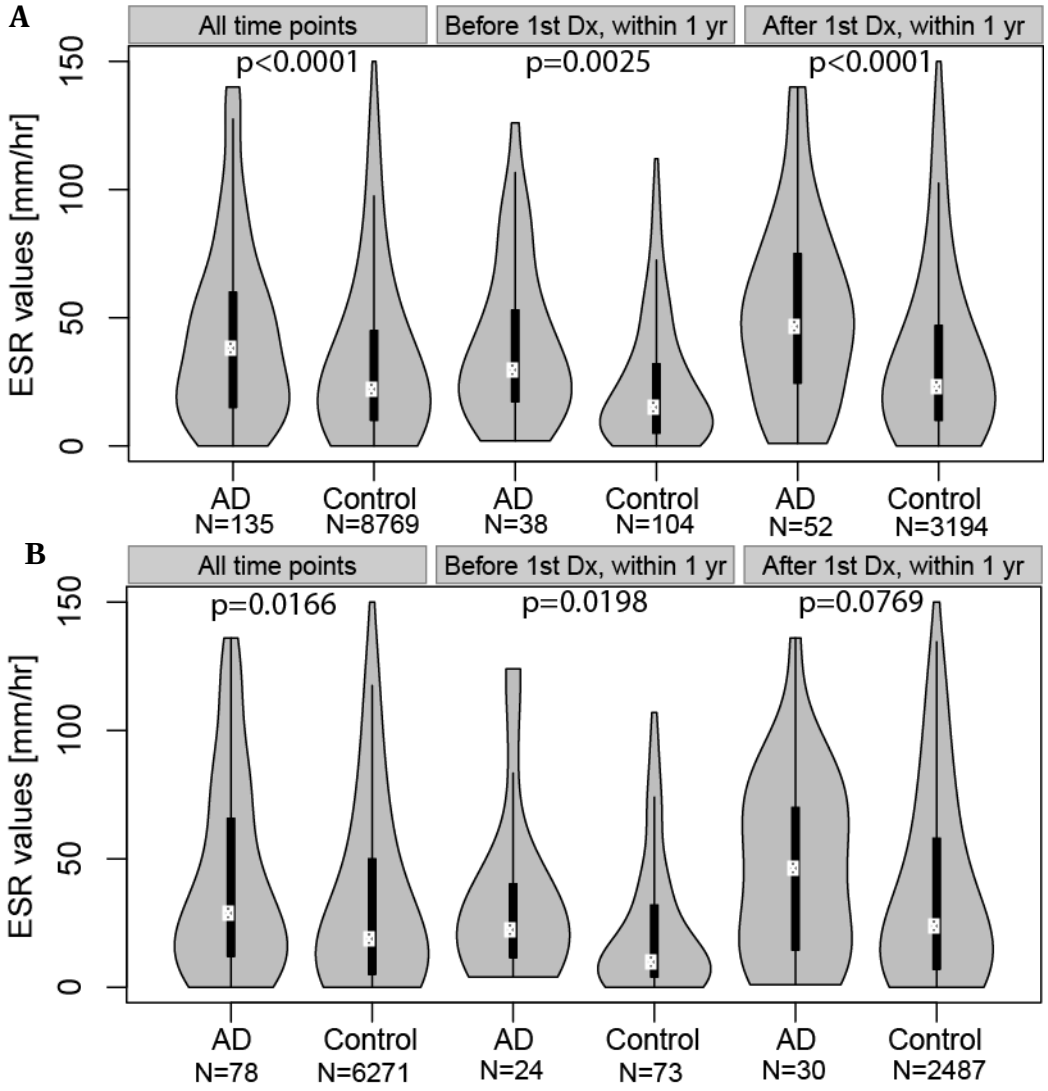
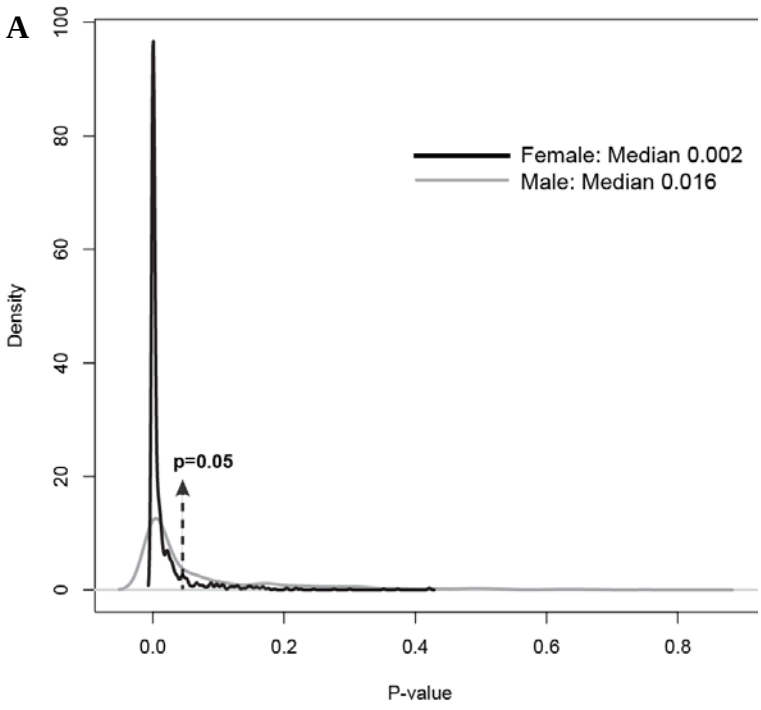


Figure 1: Violin plots (combination a boxplot and a kernel density plot) for ESR associated with AD overall time points, within 1 year lab tested prior to the 1st diagnosis, within 1 year lab tested after the 1st diagnosis for female (1A) and male (1B). In the black box plots, the bold black line boundaries indicate the 25th, 75th percentiles of ESR values, and white center squares indicate the median value of ESR. The outside grey shapes indicate density of the number of samples. P-values are reported by Mann–Whitney U test.

To match the ages of control patients to AD patients, we also performed a random sampling method to randomly select the same number of patients from controls whose ages fit the same distribution to the ages of the AD patients. As ESR is known to have values ranging from zero to higher, and with zero known to be the most frequently resulted normal value, we calculated a one-side p-value from T test by evaluating whether the mean of the lab value is higher in the AD

patients, and repeated the process 1,000 times to generate the p-value distribution. We again tested for ESR values measured within 1 year prior to the 1st diagnosis and after the 1st diagnosis using the random sampling method to match the ages between control and AD cohorts. For instance, we randomly selected 38 female control patients matching the ages in our female AD cohort, where both cohorts had a measurement of ESR within 1 year prior to the 1st diagnosis. For within 1 year lab tested prior to the 1st diagnosis, the median p-values are 0.002 for female and 0.016 for male. For within 1 year lab tested after the 1st diagnosis, the median p-values are 0.025 for female and 0.161 for male. The distributions of p-values for prior to the 1st diagnosis and after the 1st diagnosis were shown in Figure 2A and Figure 2B. With the ESR being higher in AD cohorts compared to selected age-matched controls, this suggests that ESR might not be significantly confounded by age in our study.



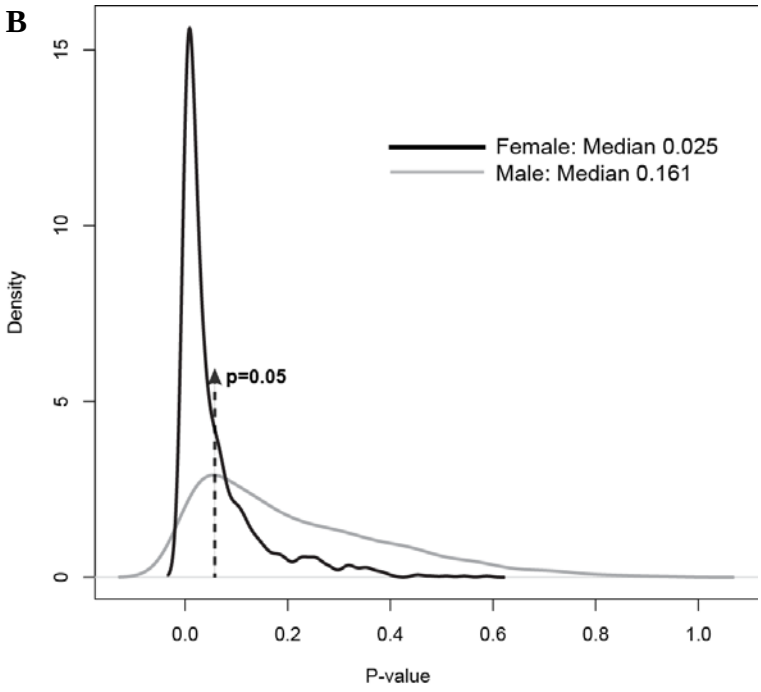


Figure 2: P-value distribution comparing age-matched AD and control groups for female (black curve) and male (grey curve) with 1,000 random samplings. ESR lab values within 1 year prior to our 1st diagnosis (2A), and ESR lab values within 1 year after our 1st diagnosis (2B). Dash line with arrow indicates p-value = 0.05.

4. Discussion

We developed a systematic approach to identify genetic associations between traits and diseases susceptibility based on common genetic architecture, aiming at identifying potential novel prognostic or risk markers for disease. In this study we focused on traits associated with Alzheimer's disease (AD) as a proof of concept, and we identified 6 clinical traits associated with AD. Five of these traits were known but one was a novel finding. We retrospectively validated our novel finding using EMR data from more than 15,000 patients at SHC.

We observed a significant association between ESR and AD, especially in female patients above 50 years old. Female patients who had abnormally elevated ESR levels had 2.31 higher chance of developing subsequent AD within a year of that lab test, compared to control patients, indicating ESR is a risk factor to AD that could be tested in a prospective trial for AD prognosis. A previous study has also reported the increased trend for ESR in AD female, although it did not reach significance due to a very small sample size (*35*). Moreover, we found that ESR persists in its elevation in female patients diagnosed with AD, suggesting that inflammation may play a role in the pathophysiology of AD (*36*), but we cannot rule out its elevation as secondary due to therapy of AD. A possible mechanism involve the complement gene inflammatory pathway including *C3*, *C4* and *C1Q* (*27-29*) as *CR1* was in common with ESR, currently used as a non-specific inflammation marker. If ESR proves to be a useful marker in specific prospective trials, we would also suggest that patients diagnosed with AD could be closely monitored for ESR as a trigger for intervention modification, such as adjusting non-steroidal anti-inflammatory medications (*36*). We would suggest that a robust prediction model could be developed

combining ESR and other current risk factors including age, lipid panel, and environmental factors and validated using multi-center EMR data, then further validated in a prospective study.

Presently, the small sample size for the case cohort represents the limiting factor for a broader implication. We acknowledge that AD patients are relatively older than the control and the control are not exactly matched, as we used a retrospective study design based on our EMR, and not a randomized prospective trial.

Though we showed 6 significant traits with q-value ≤ 0.01, we acknowledge that threshold parameters could be altered. For example, the seventh trait on our list associated with AD would be high-density lipoprotein cholesterol (HDL-C) level, with $q = 0.011$. A recent study has shown that higher levels of HDL-C were indeed associated with a decreased risk of both probable and possible AD compared with lower HDL-C levels (*37*).We could increase our significance cutoff for more novel findings. However, in this study, we used a well-accepted stringent q-value cutoff from random shuffling to avoid identifying false positive.

We do acknowledge that our discovered association and validation cannot fully distinguish the causal direction of the association or if a single associated mutation in a shared gene systematically influences both phenotypes. Regardless, we do suggest that the strategy we adopted here captures and exploits relevant genetic association between disease and traits. The approach described here could in theory be applied to any disease in order to refine their risk factors model. Investigating clinical traits that share genetic architecture with a disease, and validating these traits through EMR data is a powerful and efficient way to identify risk factors, prognostics, and diagnostic markers for complex disease.

5 Acknowledgments

This work was supported by Lucile Packard Foundation for Children's Health. We thank Chirag Patel for discussion and suggestions to this work. We thank Alex Skrenchuk for computer cluster IT support.

References

1. A.M. Minino *et al.*, *Natl Vital Stat Rep* **59**, 1 (Dec 7, 2011).
2. L.E. Hebert *et al.*, *Arch Neurol* **60**, 1119 (Aug, 2003).
3. J. Povova *et al.*, *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* **156**, 108 (Jun, 2012).
4. K. Andersen *et al.*, *Neurology* **53**, 1992 (Dec 10, 1999).
5. K. Blennow *et al.*, *Lancet* **368**, 387 (Jul 29, 2006).
6. T.D. Bird. in *GeneReviews* (eds. Pagon, R.A. *et al.*) (Seattle (WA), 1993).
7. L. Fratiglioni *et al.*, *Ann Neurol* **33**, 258 (Mar, 1993).
8. S.T. Pendlebury *et al.*, *Lancet Neurol* **8**, 1006 (Nov, 2009).
9. R.A. Whitmer *et al.*, *Neurology* **71**, 1057 (Sep 30, 2008).
10. A. Solomon *et al.*, *Dement Geriatr Cogn Disord* **28**, 75 2009).
11. H.C. Hendrie *et al.*, *Alzheimers Dement* **2**, 12 (Jan, 2006).
12. M. Kivipelto *et al.*, *Arch Neurol* **62**, 1556 (Oct, 2005).
13. T.M. Sivanandam *et al.*, *Neurosci Biobehav Rev* **36**, 1376 (May, 2012).
14. , *Nature* **447**, 661 (Jun 7, 2007).
15. A.D. Johnson *et al.*, *BMC Med Genet* **10**, 6 2009).
16. U.P. Steinbrecher *et al.*, *Arterioscler Thromb* **12**, 608 (May, 1992).
17. K.I. Goh *et al.*, *Proc Natl Acad Sci U S A* **104**, 8685 (May 22, 2007).
18. H.J. Lowe *et al.*, *AMIA Annu Symp Proc* **2009**, 391 2009).
19. R. Chen *et al.*, *PLoS One* **5**, e13574 2010).
20. R. Chen *et al.*, *PLoS Genet* **8**, e1002621 (Apr, 2012).
21. C.J. Patel *et al.*, *Bioinformatics* **28**, i121 (Jun 15, 2012).
22. S. Suthram *et al.*, *PLoS Comput Biol* **6**, e1000662 (Feb, 2010).
23. R. Chen *et al.*, *Nat Methods* **4**, 879 (Nov, 2007).
24. H.C. Wu *et al.*, *Acm Transactions on Information Systems* **26**, 2008).
25. J.D. Storey *et al.*, *Proc Natl Acad Sci U S A* **100**, 9440 (Aug 5, 2003).
26. R. Ihaka *et al.*, *Journal of computational and graphical statistics* **5**, 299 1996).
27. S.K. Nadar *et al.*, *J Hum Hypertens* **21**, 261 (Apr, 2007).
28. A. Muscari *et al.*, *Am J Med* **98**, 357 (Apr, 1995).
29. M. Naesens *et al.*, *J Am Soc Nephrol* **20**, 1839 (Aug, 2009).
30. M. van Oijen *et al.*, *Ann Neurol* **59**, 139 (Jan, 2006).
31. S.E. O'Bryant *et al.*, *J Geriatr Psychiatry Neurol* **23**, 49 (Mar, 2010).
32. P. Caramelli *et al.*, *Acta Neurol Scand* **100**, 61 (Jul, 1999).
33. T. Matsuzaki *et al.*, *Neurology* **77**, 1068 (Sep 13, 2011).
34. G.T. Lesser *et al.*, *Dement Geriatr Cogn Disord* **27**, 42 2009).
35. D. Robinson *et al.*, *Journal of the American Geriatrics Society* **43**, 1177 1995).
36. E.E. Tuppo *et al.*, *Int J Biochem Cell Biol* **37**, 289 (Feb, 2005).
37. C. Reitz *et al.*, *Arch Neurol* **67**, 1491 (Dec, 2010).