# Integration of bioinformatics and imaging informatics for identifying rare PSEN1 variants in Alzheimer's disease

## Citation

## Published Version

## Permanent link

## Terms of Use

# Share Your Story

BMC Medical Genomics

CrossMark

# Integration of bioinformatics and imaging informatics for identifying rare *PSEN1* variants in Alzheimer's disease

Kwangsik Nho[1,3,13*], Emrin Horgusluoglu[2,13], Sungeun Kim[1,3,13], Shannon L. Risacher[1,13], Dokyoon Kim[4], Tatiana Foroud[1,2,3,13], Paul S. Aisen[6], Ronald C. Petersen[7], Clifford R. Jack Jr.[8], Leslie M. Shaw[5], John Q. Trojanowski[5], Michael W. Weiner[9,10], Robert C. Green[11], Arthur W. Toga[12], Andrew J. Saykin[1,2,3,13*] and ADNI

## Abstract

**Background:** Pathogenic mutations in *PSEN1* are known to cause familial early-onset Alzheimer's disease (EOAD) but common variants in *PSEN1* have not been found to strongly influence late-onset AD (LOAD). The association of rare variants in *PSEN1* with LOAD-related endophenotypes has received little attention. In this study, we performed a rare variant association analysis of *PSEN1* with quantitative biomarkers of LOAD using whole genome sequencing (WGS) by integrating bioinformatics and imaging informatics.

**Methods:** A WGS data set ($N = 815$) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort was used in this analysis. 757 non-Hispanic Caucasian participants underwent WGS from a blood sample and high resolution T1-weighted structural MRI at baseline. An automated MRI analysis technique (FreeSurfer) was used to measure cortical thickness and volume of neuroanatomical structures. We assessed imaging and cerebrospinal fluid (CSF) biomarkers as LOAD-related quantitative endophenotypes. Single variant analyses were performed using PLINK and gene-based analyses of rare variants were performed using the optimal Sequence Kernel Association Test (SKAT-O).

**Results:** A total of 839 rare variants (MAF $< 1/\sqrt{(2\ N)} = 0.0257$) were found within a region of ±10 kb from *PSEN1*. Among them, six exonic (three non-synonymous) variants were observed. A single variant association analysis showed that the *PSEN1* p. E318G variant increases the risk of LOAD only in participants carrying *APOE* ε4 allele where individuals carrying the minor allele of this *PSEN1* risk variant have lower CSF Aβ$_{1-42}$ and higher CSF tau. A gene-based analysis resulted in a significant association of rare but not common (MAF ≥ 0.0257) *PSEN1* variants with bilateral entorhinal cortical thickness.

**Conclusions:** This is the first study to show that *PSEN1* rare variants collectively show a significant association with the brain atrophy in regions preferentially affected by LOAD, providing further support for a role of *PSEN1* in LOAD. The *PSEN1* p. E318G variant increases the risk of LOAD only in *APOE* ε4 carriers. Integrating bioinformatics with imaging informatics for identification of rare variants could help explain the missing heritability in LOAD.

**Keywords:** Whole genome sequencing, Imaging genetics, Gene-based association of rare variants, *PSEN1*

---

* Correspondence: knho@iupui.edu; asaykin@iupui.edu
[1]Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA
Full list of author information is available at the end of the article

BioMed Central

Nho *et al. BMC Medical Genomics* 2016, **9**(Suppl 1):30

Page 12 of 97

## Background

Late-onset Alzheimer's disease (LOAD) is a progressive neurodegenerative condition with no validated disease modifying treatment. With the heritability of LOAD estimated to be as high as 80 %, a better understanding of the genetic susceptibility factors of LOAD would advance strategies for early detection and treatment [1, 2]. Recent large-scale genome-wide association studies (GWAS) have identified and confirmed approximately twenty-one LOAD-associated genes in addition to *APOE*, whose ε4 allele is the best established and the most significant genetic risk factor [3]. While about 50 % of LOAD heritability is accounted for by all of the known LOAD susceptibility genes including *APOE*, a substantial proportion of the heritability for LOAD remains to be identified [1]. A growing body of evidence highlighting the role of rare variants has opened exciting avenues for discovering novel genetic factors to explain some of the missing heritability and facilitate a comprehensive understanding of LOAD.

Rapid advancement of next generation sequencing technologies has facilitated the search for genetic susceptibility factors that influence disease risk and become a key technique for detecting pathogenic variants in human diseases [4]. Our understanding of the impact of the genetic variation on human diseases has been greatly advanced using high-throughput sequencing [5]. Whole genome sequencing (WGS) has been used to obtain the most comprehensive genetic variation of an individual and perform detailed evaluation of all genetic variation [6]. Several sequencing-based association studies could identify functional risk variants with large effects on LOAD pathogenesis within *TREM2*, *ABCA7*, and *PLD3* genes [7–11].

Pathogenic mutations in *PSEN1* are known to cause familial early-onset Alzheimer's disease (EOAD) but common variants in *PSEN1* have not been found to strongly influence LOAD [12]. Thus, the association of rare variants in *PSEN1* with LOAD-related endophenotypes has received little attention.

Accumulating evidence shows that common and rare risk variants are likely to co-exist at the same locus (known as pleomorphic risk loci) [13]. Deep re-sequencing-based association studies could identify functional risk variants within known susceptibility genes such as *ABCA7* [10]. In this study, we performed a rare variant association analysis of *PSEN1* with quantitative biomarkers of LOAD using WGS. Integration of bioinformatics and imaging informatics will provide a comprehensive and integrative approach to identifying a LOAD-specific genetic variation. In particular, imaging genetics combines neuroimaging such as MRI and PET with genetics for studying the influence of genetic variation on brain structure and function [14]. Quantitative endophenotypes increase detection power for

rare variant association analysis and give additional information to interpret the association of variants by suggesting potential biological mechanisms by which the identified variants could lead to disease [14].

## Methods

### Subjects

All individuals included in these analyses were participants in the Alzheimer's Disease Neuroimaging Initiative Phase 1 (ADNI-1) and its subsequent extensions (ADNI-GO/2). The initial phase (ADNI-1) was launched in 2003 to test whether serial magnetic resonance imaging (MRI), position emission tomography (PET), other biological markers, and clinical and neuropsychological assessment could be combined to measure the progression of MCI and early AD. The ADNI-1 participants were recruited from 59 sites across the U.S. and Canada and include approximately 200 cognitively normal older individuals (healthy controls (HC)), 400 patients diagnosed with MCI, and 200 patients diagnosed with early probable AD aged 55–90 years. ADNI-1 has been extended to its subsequent phases (ADNI-GO and ADNI-2) for follow-up for existing participants and additional new enrollments. Inclusion and exclusion criteria, clinical and neuroimaging protocols, and other information about ADNI have been published previously [15] and can be found at www.adni-info.org. Demographic information, raw scan data, *APOE* and whole genome sequencing data, neuropsychological test scores, and diagnostic information are available from the ADNI data repository (http://www.loni.usc.edu/ADNI/). Written informed consent was obtained at the time of enrollment for imaging and genetic sample collection and protocols of consent forms were approved by each participating sites' Institutional Review Board (IRB).

### Whole genome sequencing (WGS) analysis

WGS was performed on blood-derived genomic DNA samples obtained from 817 ADNI participants. Samples were sequenced on the Illumina HiSeq2000 using paired-end read chemistry and read lengths of 100 bp (www.illumina.com). The resulting Illumina qseq files were converted into fastq files, a text-based format for storing both sequence reads and their corresponding quality information in Phred format. Quality checks and read statistics are performed on raw sequence data in FASTQ format using FastQC. Short-read sequences are mapped to the NCBI reference human genome (build 37.72) using BWA, allowing for up to two mismatches in each read [16]. During the alignment, we use only bases with Phred Quality > 15 in each read to include soft clipping of low-quality bases, retain only uniquely mapped pair-end reads, and remove potential PCR duplicates. After completing initial alignment, the

Nho *et al. BMC Medical Genomics* 2016, **9**(Suppl 1):30

Page 13 of 97

alignment is further refined by locally realigning any suspicious reads. The reported base calling quality scores obtained from the sequencer are re-calibrated to account for covariates of base errors such as sequencing technology and machine cycle [17]. Finally, the realigned reads are written to a BAM file for further analysis. The analysis-ready BAM files are analyzed to identify all variants with statistical evidence for an alternate allele present among samples using GATK HaplotypeCaller for multi-sample variant callings [17]. For variants which pass recommended variation quality criteria, ANNOVAR is used to annotate all variants (SNPs (single nucleotide polymorphism) and short insertion/deletions (indels)) [18]. We performed standard quality control procedures in WGS to assess the quality of WGS and to remove individuals and genetic variants with poor quality. We excluded variants that did not pass the variant quality score recalibration step using GATK in the WGS analysis pipeline and we removed variants whose genotype quality (GQ) scores < 20. The quality of the variant calls was assessed by comparing sequencing-derived SNPs with those obtained from the Illumina Omni 2.5 M genotyping array in order to estimate the concordance rate for each individual. Among 817 subjects, two subjects had concordance rates less than 99 % and had been removed from our analysis. The remaining subjects had a mean concordance rate of 99.9 % (Fig. 1).
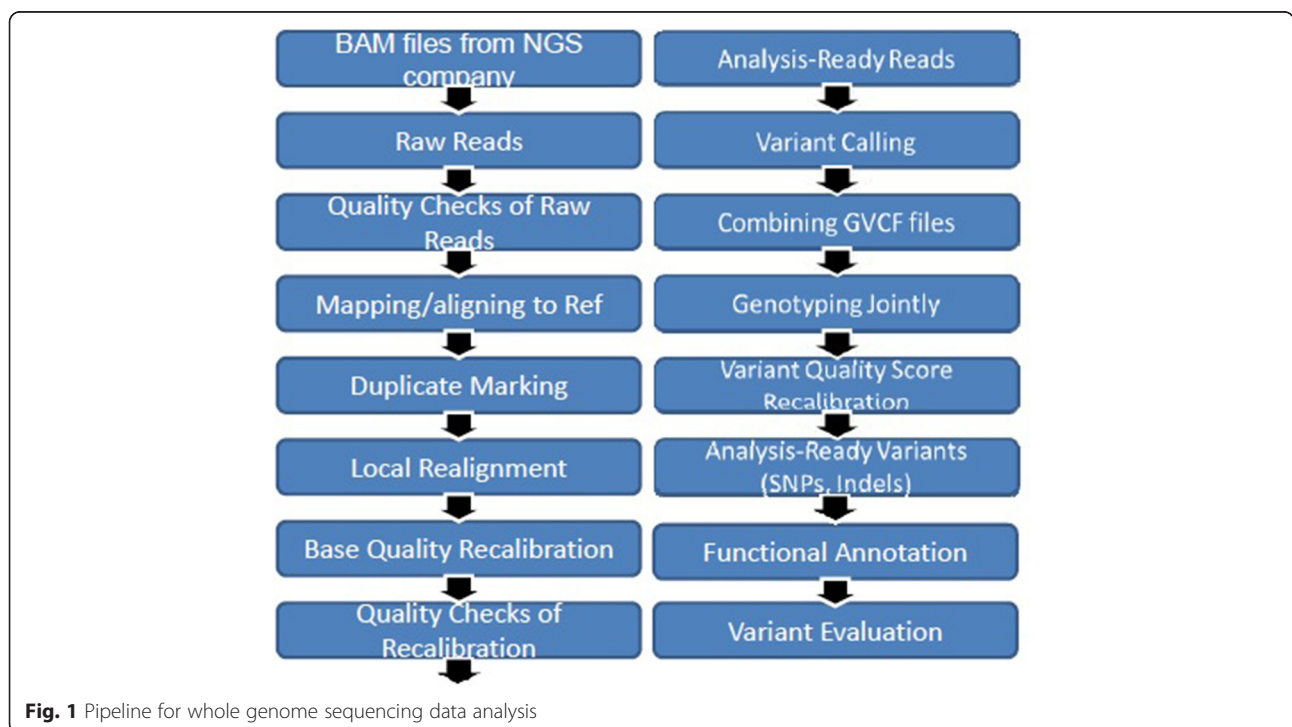
## Subject selection

Since population stratification is known to cause spurious association in disease studies, we restricted our analyses to only subjects that clustered with CEU (Utah residents with Northern and Western European ancestry from the CEPH collection) + TSI (Toscani in Italia) populations using HapMap 3 genotype data and the multidimensional scaling (MDS) analysis (www.hapmap.org) [19].

## CSF measurements

Baseline CSF samples were obtained using previously reported methods for 3 CSF measurements (Amyloid-β 1–42 peptide ($A\beta_{1-41}$), total tau (t-tau), and tau phosphorylated at the threonine 181 (p-tau$_{181p}$)) as described [20]. Subjects who had at least one value greater or smaller than 4 SD (standard deviation) from the mean value of each of 3 CSF variables were removed from the analysis as extreme outliers [20].

## Imaging processing

$T_1$-weighted brain MRI scans at baseline were acquired using a sagittal 3D MP-RAGE sequence following the ADNI MRI protocol [21]. As detailed in previous studies [22], a widely employed automated MRI analysis technique was used to process MRI scans: FreeSurfer V5.1 software (http://surfer.nmr.mgh.harvard.edu/). FreeSurfer was used to process and extract brain-wide target MRI imaging phenotypes (region volume and cortical thickness)



**Fig. 1** Pipeline for whole genome sequencing data analysis

Nho et al. BMC Medical Genomics 2016, **9**(Suppl 1):30

Page 14 of 97

by automated segmentation and parcellation. The cortical surface was reconstructed to measure thickness at each vertex on surface. The cortical thickness was calculated by taking the Euclidean distance between the grey/white boundary and the grey/cerebrospinal fluid (CSF) boundary at each vertex on surface. For surface-based comparison of the cortical thickness, all individual cortical surfaces were registered to a common surface template, which was an average created from all cognitively normal control subjects. The cortical thickness was smoothed with 10 mm FWHM Gaussian kernel to improve the signal-to-noise ratio and statistical power.

### Statistical analysis

Using all WGS-identified SNPs in the *PSEN1* gene region, we performed a gene-based analysis of rare variants with the optimal Sequence Kernel Association Test (SKAT-O) [23]. Rare variants were defined as variants with less than $1/\sqrt{(2\ N)} = 0.0257$ minor allele frequency (MAF) in our WGS sample. SKAT-O, which is an optimal unified approach for testing the association between rare variants and phenotypes in sequencing association studies and allows for heterogeneous effect of all variants within each gene, performed a score test for the model that includes the variants within each gene. Burden tests collapse information for multiple rare variants into a single genetic score and test for association of the score with a trait of interest [24]. In this study, we used the Morris and Zeggini (MZ) test as a burden test [25]. The MZ test uses a dominant genetic model to compute genetic scores as the number of rare variants for which an individual carries at least one copy of the minor allele [24]. The burden tests are powerful when a large proportion of variants are causal and effects are in the same direction. We performed association analysis using all SNPs. A single rare variant association analysis was tested using linear regression under a dominant genetic model in PLINK. Potential confounding factors (age, gender, years of education, intracranial volume (ICV), and MRI field strength) were used as covariates.

**Table 1** Demographic characteristics of study participants

|  | HC | EMCI | LMCI | AD |
|---|---|---|---|---|
| N | 259 | 219 | 232 | 47 |
| Gender (M/F) | 132/127 | 121/98 | 148/84 | 30/17 |
| Age | 74.3 (5.5) | 71.1 (7.4) | 73.2 (7.3) | 75.2 (9.3) |
| Education | 16.5 (2.7) | 16.0 (2.7) | 16.1 (3.0) | 15.7 (2.7) |
| MMSE | 29.1 (1.2) | 28.4 (1.5) | 27.5 (1.7) | 22.9 (2.0) |
| *APOE* (ε4−/ε4+) | 189/70 | 132/87 | 113/119 | 14/33 |

Values given are means (standard deviation), *MMSE* mini-mental state examination, *HC* cognitive normal older adult, *EMCI* early mild cognitive impairment (MCI), *LMCI* late MCI, *AD* Alzheimer's disease

**Table 2** Single nucleotide polymorphisms (SNP) within *PSEN1* ± 10 kilobase (kb)

| Common SNP (MAF ≥ 0.0257) | Novel | 5 | intronic: 4; downstream: 1 |
|---|---|---|---|
|  | Known | 181 | intergenic: 35; intronic: 141; 3' UTR: 7; up (down) stream: 3 |
| Rare SNP (MAF < 0.0257) | Novel | 498 | intergenic: 99; intronic: 372; exonic: 2; 3' UTR: 15; 5' UTR: 1; up (down) stream: 9 |
|  | Known | 341 | intergenic: 57; intronic: 257; exonic: 4; 3' UTR: 19; 5'UTR: 5; up (down) stream: 3 |

## Results

A total of 757 participants met criteria (quality controls and population stratification) for inclusion in analysis. There were 47 patients with Alzheimer's disease (AD), 219 patients with early mild cognitive impairment (MCI), 232 patients with late MCI, and 259 cognitively normal older adults (HC) (Table 1). As expected, AD patients were found to have significantly lower Mini-Mental State Examination (MMSE) scores. Furthermore, the *APOE* ε4 allele frequency was significantly higher in patients with AD.

### Sequencing of the PSEN1 gene region

From an established WGS analysis pipeline, we found a total of 1,025 SNPs within a region of ±10 kb from *PSEN1*. Among 1,025 variants, there are 186 common (minor allele frequency (MAF) ≥ 0.0257) and 839 rare SNPs (MAF < 0.0257). Of 6 exonic rare variants, we found 3 nonsynonymous SNPs (Table 2).

### Association of a PSEN1 rare variant p. E318G with CSF Biomarkers

To replicate the previous finding the risk variant (p. E318G) in *PSEN1* is significantly associated with LOAD-specific biomarkers, we performed an association using CSF biomarkers $A\beta_{1-42}$ and t-tau, biomarkers of LOAD-associated pathologic changes in the brain. 583 participants with WGS data also had CSF biomarkers. A single variant association analysis showed that the *PSEN1* p. E318G variant was significantly associated with CSF biomarkers and increased the risk of LOAD in participants with *APOE* ε4 allele ($p < 0.05$; Table 3). Individuals carrying the minor allele of the *PSEN1* risk variant have lower CSF $A\beta_{1-42}$ and higher CSF t-tau. However, the *PSEN1* p. E318G variant was not significantly associated with CSF biomarkers in participant group not carrying *APOE* ε4 allele. The percentage of participants without

**Table 3** Association results (*p*-values) of *PSEN1* p. E318G variant for quantitative trait analysis using a dominant model

|  | All participants (N = 583) | Participants with *APOE* ε4 (N = 234) | Participants without *APOE* ε4 (N = 349) |
|---|---|---|---|
| $A\beta_{1-42}$ | 0.6199 | 0.0254 | 0.2733 |
| t-tau | 0.3795 | 0.0125 | 0.8673 |

Nho *et al. BMC Medical Genomics* 2016, **9**(Suppl 1):30

Page 15 of 97

*APOE* ε4 allele having the minor allele (G) of *PSEN1* p. E318G is 6 % (21 participants), which is two times higher than that of participants with *APOE* ε4 allele (3 %; 7 participants) having the minor allele of *PSEN1* p. E318G. Furthermore, the *PSEN1* p. E318G variant was not significantly associated with CSF biomarkers in all participants regardless of the *APOE* ε4 status.

### Gene-based association of PSEN1 rare variants with LOAD-specific imaging biomarkers

To test the hypothesis that *PSEN1* rare variants would be associated with structural changes in LOAD-related brain regions, we assessed entorhinal cortical thickness (EntCtx) as LOAD-related quantitative endophenotypes based on prior studies indicating that regional structural brain change in LOAD occurs initially and most severely in the entorhinal cortex and hippocampus before spreading throughout the entire brain [26]. Of the 755 of the 757 participants with WGS data, had usable high resolution T1-weighted structural MRI at baseline.

Gene-based association analysis of rare SNPs in the *PSEN1* gene with both SKAT-O and a burden test resulted that *PSEN1* rare variants were significantly associated with entorhinal cortical thickness after correction for multiple comparisons. The significant association was increased after adjustment for the *APOE* ε4 status ($p < 0.05$; Table 4).

**Table 4** Gene-based association results (*p*-values) for imaging biomarkers (entorhinal cortical thickness (EntCtx)) using rare variants (MAF < 0.0257)

| | Burden | SKAT-O | SKAT-O after adjusting for *APOE* ε4 status |
|---|---|---|---|
| Left EntCtx | 0.009 | 0.015 | 0.010 |
| Right EntCtx | 0.027 | 0.046 | 0.032 |

To examine the cortical topography of all rare variants in *PSEN1*, an unbiased whole-brain multivariate analysis of cortical thickness was performed on a vertex-by-vertex basis to detect additional regions of association. We collapsed all rare variants into single risk score by combining minor allele counts into a single risk score with a dominant genetic model. A general linear model was constructed using age, gender, year of education, MRI field strength, and ICV as covariates. A random field theory based adjustment was used to correct for multiple comparisons to retain a 0.05 level of significance. Figure 2 displays the results of the main effect of all rare variants after adjusting for *APOE* ε4 status. Highly significant clusters associated with the risk score were found in bilateral temporal, bilateral parietal, and right frontal lobes regions (Table 5), where subjects having high risk scores showed thinner mean
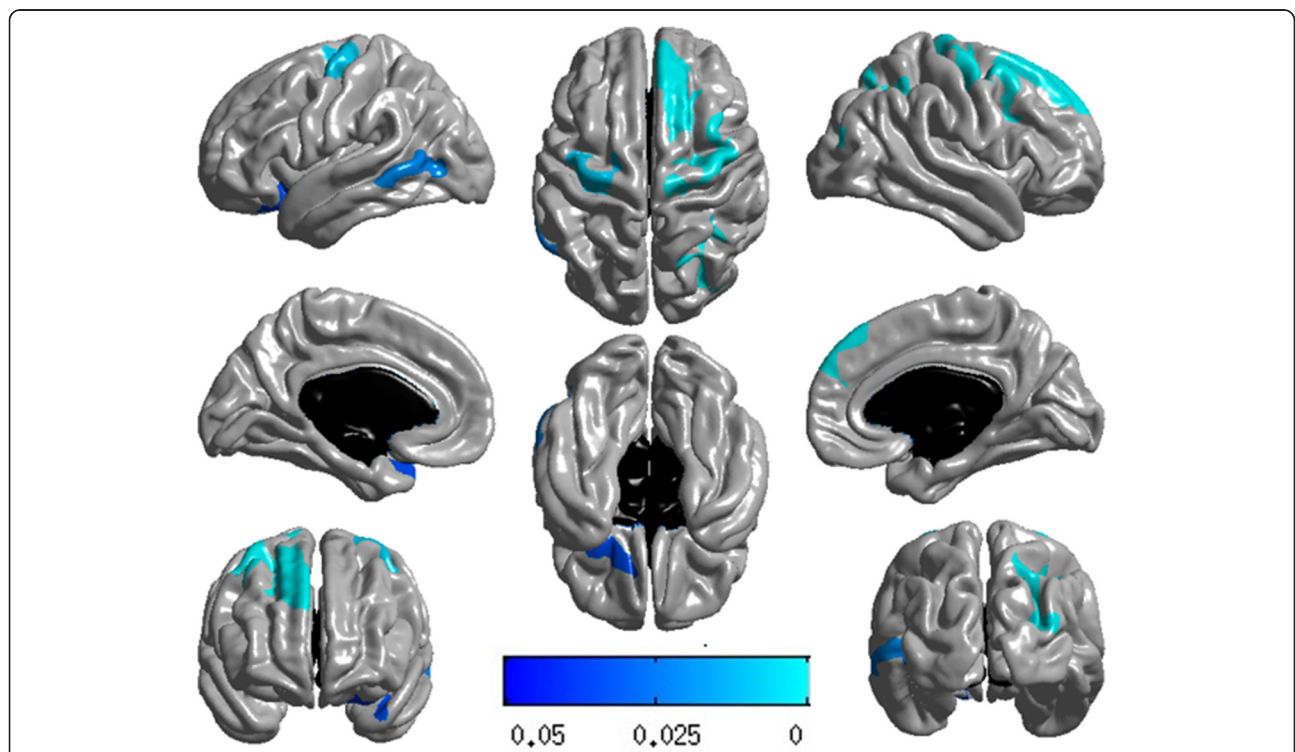


**Fig. 2** Surface-based whole-brain analysis results. A whole-brain multivariate analysis of cortical thickness was performed on a vertex-by-vertex basis to visualize the topography of genetic association in an unbiased manner. Statistical maps were thresholded using a random field theory adjustment to a corrected significance level of $p = 0.05$

Nho *et al. BMC Medical Genomics* 2016, **9**(Suppl 1):30

Page 16 of 97

cortical thickness compared with the participants having lower risk scores (Fig. 2).

### Association of common SNPs in the PSEN1 gene with LOAD-specific imaging biomarkers

A single variant and a gene-based association analyses of common SNPs in the *PSEN1* gene identified no significant association with entorhinal cortical thickness passed a multiple comparison correction (data not shown).

### Discussion and Conclusions

An association analysis of rare variants was performed on 757 ADNI WGS samples to investigate the influence of genetic variation in the *PSEN1* gene on LOAD-related imaging biomarkers. To our knowledge, this is the first study to show that *PSEN1* rare variants collectively show a significant association with the brain atrophy in regions preferentially affected by LOAD using integrated informatics methodologies. Our results indicated that rare variants in *PSEN1* were significantly associated with cortical thickness after correction for multiple comparisons and the significant association was increased after adjustment for the *APOE* ε4 status.

*PSEN1* regulates APP processing by affecting on gamma secretase enzyme which cleaves amyloid precursor protein (APP) and regulates amyloid-β accumulation which is a pathological hallmark of AD [27]. In addition, *PSEN1* plays an important role in Notch signaling pathway through the cleavage of the Notch receptor and Wnt signaling pathway [28, 29]. *PSEN1* locates on chromosome 14 and mutations in this gene are autosomal dominant and cause the early-onset AD [30]. Large-scale genome-wide association study (GWAS) showed that none of SNPs at the *PSEN1* locus reached to genome-wide significance [31]. Association studies investigating rare coding variants on the gene showed that rare variants in *PSEN1* were associated with sporadic LOAD [32, 33] and *PSEN1* p. E318G variant increased the risk of LOAD only in participants carrying *APOE* ε4 allele [12].

**Table 5** Clusters significantly associated with polygenic risk scores of all rare variants within PSEN1 after multiple testing correction

| Cluster | Cluster size (# of vertex) | Anatomic localization | Corrected *p*-value |
|---|---|---|---|
| 1 | 3,798 | right frontal lobe | $8.96 \times 10^{-6}$ |
| 2 | 4,247 | right frontal lobe | $2.29 \times 10^{-5}$ |
| 3 | 3,436 | right parietal lobe | $1.23 \times 10^{-4}$ |
| 4 | 2,718 | left frontal lobe | $7.11 \times 10^{-3}$ |
| 5 | 1,915 | left temporal lobe | $2.04 \times 10^{-2}$ |
| 6 | 2,078 | left temporal lobe | $3.33 \times 10^{-2}$ |

Performing advanced sequencing data analysis (bioinformatics) and human brain imaging analysis (imaging informatics) in an integrated approach enables us to identify blood-based biomarkers for risk or protection of LOAD, leading to an improved early diagnosis and prognosis, using LOAD-specific endophenotypes. Furthermore, use of quantitative endophenotypes substantially increase detection power for rare variant association analysis and holds great promise for discovery of variation mechanically related to AD pathophysiology. Confirmation of our results in independent and larger cohorts will be warranted.

In conclusion, we used whole genome sequencing to perform an association analysis of rare variants in *PSEN1* with LOAD-related imaging biomarkers. Our results illustrate the potential of integration of informatics methodologies to identify novel diagnostic/therapeutic targets for LOAD and understand the genetics and pathobiology of LOAD.

Nho *et al. BMC Medical Genomics* 2016, **9**(Suppl 1):30

Page 17 of 97

## Availability of data and materials
Demographic information, raw neuroimaging scan data, *APOE* and whole genome sequencing data, neuropsychological test scores, and diagnostic information are available from the ADNI data repository (http://www.loni.usc.edu/ADNI/).

## Authors' contributions
All authors contributed substantively to this work. KN, EH, SK, SLR, and AJS were involved in study conception and design. KN, EH, DK, and AJS were involved in data organization, whole genome sequencing analysis and statistical analyses. TF, PSA, RCP, RCG, CRJ, LMS, JQT, RCG, AWT, MWW and AJS were involved in coordination and data collection and processing for ADNI. KN and AJS drafted the report and prepared all figures and tables. All authors were involved in reviewing and editing of the manuscript. All authors read and approved the final manuscript.

## Competing interests
The authors declare that they have no competing interests.

## Consent for publication
Not applicable.

## Ethics approval and consent to participate
Written informed consent was obtained at the time of enrollment for imaging and genetic sample collection and protocols of consent forms were approved by each participating sites' Institutional Review Board (IRB).

## Author details
[1]Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA. [2]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. [3]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA. [4]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, USA. [5]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA. [6]Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA. [7]Department of Neurology, Mayo Clinic Minnesota, Rochester, MN, USA. [8]Department of Radiology, Mayo Clinic Minnesota, Rochester, MN, USA. [9]Departments of Radiology, Medicine, and Psychiatry, University of California-San Francisco, San Francisco, CA, USA. [10]Department of Veterans Affairs Medical Center, San Francisco, CA, USA. [11]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [12]The Institute for Neuroimaging and Informatics and Laboratory of Neuro Imaging, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA. [13]Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA.

## References
1. Ridge PG, Mukherjee S, Crane PK, Kauwe JS, Alzheimer's Disease Genetics C. Alzheimer's disease: analyzing the missing heritability. PLoS ONE. 2013;8: e79771.
2. Gatz M, Pedersen NL, Berg S, Johansson B, Johansson K, et al. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. J Gerontol A Biol Sci Med Sci. 1997;52:M117–25.
3. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45:1452–8.
4. Esteban-Jurado C, Vila-Casadesus M, Garre P, Lozano JJ, Pristoupilova A, et al. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. Genet Med. 2015;17:131–42.
5. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30:418–26.
6. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet. 2010;11:415–25.
7. Guerreiro R, Hardy J. TREM2 and neurodegenerative disease. N Engl J Med. 2013;369:1569–70.
8. Guerreiro RJ, Lohmann E, Bras JM, Gibbs JR, Rohrer JD, et al. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. JAMA Neurol. 2013;70:78–84.
9. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013;368:117–27.
10. Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat Genet. 2015;47:445–7.
11. Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature. 2014;505:550–4.
12. Benitez BA, Karch CM, Cai Y, Jin SC, Cooper B, et al. The PSEN1, p. E318G variant increases the risk of Alzheimer's disease in APOE-epsilon4 carriers. PLoS Genet. 2013;9:e1003685.
13. Singleton A, Hardy J. A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. Hum Mol Genet. 2011;20:R158–62.
14. Medland SE, Jahanshad N, Neale BM, Thompson PM. Whole-genome analyses of whole-brain data: working within an expanded search space. Nat Neurosci. 2014;17:791–800.
15. Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, et al. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimers Dement. 2010;6:265–73.
16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
17. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.
18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
19. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet. 2010;11:459–63.
20. Kim S, Swaminathan S, Shen L, Risacher SL, Nho K, et al. Genome-wide association study of CSF biomarkers Abeta1-42, t-tau, and p-tau181p in the ADNI cohort. Neurology. 2011;76:69–79.
21. Jack Jr CR, Bernstein MA, Fox NC, Thompson P, Alexander G, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging. 2008;27:685–91.
22. Risacher SL, Shen L, West JD, Kim S, McDonald BC, et al. Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. Neurobiol Aging. 2010;31:1401–18.
23. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012;91:224–37.
24. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95:5–23.
25. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2010;34:188–93.
26. Scahill RI, Schott JM, Stevens JM, Rossor MN, Fox NC. Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. Proc Natl Acad Sci U S A. 2002;99:4703–7.
27. Duff K, Eckman C, Zehr C, Yu X, Prada CM, et al. Increased amyloid-beta42 (43) in brains of mice expressing mutant presenilin 1. Nature. 1996;383:710–3.
28. De Strooper B, Saftig P, Craessaerts K, Vanderstichele H, Guhde G, et al. Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. Nature. 1998;391:387–90.
29. Zhang C, Wu B, Beglopoulos V, Wines-Samuelson M, Zhang D, et al. Presenilins are essential for regulating neurotransmitter release. Nature. 2009;460:632–6.
30. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature. 1995;375:754–60.
31. Gerrish A, Russo G, Richards A, Moskvina V, Ivanov D, et al. The role of variation at AbetaPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease. J Alzheimers Dis. 2012;28:377–87.

Nho *et al. BMC Medical Genomics* 2016, **9**(Suppl 1):30

Page 18 of 97

32. Sassi C, Guerreiro R, Gibbs R, Ding J, Lupton MK, et al. Investigating the role of rare coding variability in Mendelian dementia genes (APP, PSEN1, PSEN2, GRN, MAPT, and PRNP) in late-onset Alzheimer's disease. Neurobiol Aging. 2014;35(2881):e2881–6.
33. Cruchaga C, Haller G, Chakraverty S, Mayo K, Vallania FL, et al. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. PLoS ONE. 2012;7:e31039.