

RESEARCH

Open Access



A novel approach for multi-SNP GWAS and its application in Alzheimer's disease

Paul M. Bodily^{1*}, M. Stanley Fujimoto¹, Justin T. Page², Mark J. Clement¹, Mark T. W. Ebbert², Perry G. Ridge² and the Alzheimer's Disease Neuroimaging Initiative

From 12th Annual Biotechnology and Bioinformatics Symposium (BIOT-2015)
Provo, UT, USA. 10-11 December 2015

Abstract

Background: Genome-wide association studies (GWAS) have effectively identified genetic factors for many diseases. Many diseases, including Alzheimer's disease (AD), have epistatic causes, requiring more sophisticated analyses to identify groups of variants which together affect phenotype.

Results: Based on the GWAS statistical model, we developed a multi-SNP GWAS analysis to identify pairs of variants whose common occurrence signaled the Alzheimer's disease phenotype.

Conclusions: Despite not having sufficient data to demonstrate significance, our preliminary experimentation identified a high correlation between GRIA3 and HLA-DRB5 (an AD gene). GRIA3 has not been previously reported in association with AD, but is known to play a role in learning and memory.

Keywords: GWAS, Multi-SNP GWAS, Alzheimer's disease, Epistasis

Background

Until recently linkage studies were the best approach to identify genes responsible for genetic diseases. However, linkage studies are most successful in monogenic disorders with highly penetrant variants, and were most often only possible in families. Unfortunately, the majority of genetic diseases are complex and therefore their genetic architectures could not be studied with linkage studies. Following completion of the human genome and the development of accurate SNP arrays, the Human HapMap project [1] cataloged the majority of common variants in the human genome. This catalog of SNPs facilitated the creation of genome-wide association studies (GWAS) [2]. In a GWAS, the co-occurrence of a given SNP and a phenotype are assessed. SNPs present (or absent) significantly more often in individuals with a particular phenotype or more extreme phenotype, are reported as disease markers (i.e., genomic variation correlated with a trait, but not

necessarily causative). GWAS can be used to study both quantitative and binary phenotypes, and were the first effective approach for studying the genetics of complex traits.

In 2005, Haines et al. [3] conducted the first GWAS, examining statistical significance between single SNPs and age-related macular degeneration. In the decade since, the success of this technique has been used to identify genetic factors associated with dozens of traits (e.g., coronary heart disease, type-1 diabetes, type-2 diabetes, rheumatoid arthritis, Crohn's disease, bipolar disorder, hypertension, Alzheimer's disease, and others [4]). The GWAS catalog co-curated by the National Human Genome Research Institute and the European Bioinformatics Institute contains reported associations for thousands of GWAS [5].

Unfortunately, GWAS have several limitations. First, most GWAS markers are thought to be non-functional, so while the marker may provide insights into regions of the genome important to a particular phenotype, unless the marker is functional it typically does not provide information about the specific biological mechanisms driving disease. Second, the majority of GWAS markers have only

*Correspondence: paulmbodily@gmail.com

¹ Computer Science Department, Brigham Young University, 84602-6576 Provo, UT, USA

Full list of author information is available at the end of the article

a very modest effect on risk for disease. Third, despite thousands of GWAS performed using progressively bigger datasets, collectively GWAS SNPs only explain a portion—often a small portion—of the total estimated genetic variance for a given trait [6].

A number of explanations exist for the unexplained genetic variance. One possibility is that gene-gene (i.e., epistasis) interactions are a feature of the genetic architecture of these traits. GWAS assume an additive model (i.e., SNPs confer disease risk independent of other SNPs) and therefore cannot be used to detect epistatic interactions. Numerous approaches have been attempted to identify epistasis including multifactor dimensionality reduction, regression (i.e., GWAS with interaction terms), and others [7], each with different pros and cons. In this manuscript we present a novel multi-SNP GWAS approach for identifying epistatic interactions and demonstrate its utility in Alzheimer's disease (AD).

AD is the most common cause-of-death with no effective treatments and has a rapidly increasing incidence worldwide [8]. Additionally, AD is the ideal phenotype to use to demonstrate the utility of our approach for two reasons: 1) epistasis has a role in the genetic architecture of AD [7, 9, 10], and 2) despite very large GWAS (Table 1) and the identification of several rare SNPs [11–13], a substantial portion of the genetic variance remains unexplained [14].

Methods

Dataset

This research used 802 whole genomes (279 control, 191 case, 332 unknown, and 444 males and 358 females) from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The genomes were processed by ADNI using the Burrows-Wheeler aligner (BWA) [15] and the best practices of the Genome Analysis Toolkit (GATK) [16]. Genomes were obtained from the 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 AD. For up-to-date information, see www.adni-info.org.

Single-SNP GWAS analysis

Single-SNP GWAS uses a modified form of linkage disequilibrium (LD) to infer relationships between single SNPs and observed phenotypes. In order to understand this approach, it is necessary to understand LD and how it is usually applied in genetic analyses. LD is a measure of how often two genomic features are inherited together

Table 1 Genes most highly associated with Alzheimer's disease

| Gene | dbSNP ID | Chr:Pos |
|-------------------|------------|--------------|
| APOE e2 | rs7412 | 19:45412078 |
| APOE e4 | rs429358 | 19:45411940 |
| CR1 | rs6656401 | 1:207692048 |
| BIN1 | rs6733839 | 2:127892809 |
| CD2AP | rs10948363 | 6:47487761 |
| EPHA1 | rs11771145 | 7:143110761 |
| CLU | rs9331896 | 8:27467685 |
| MS4A6A | rs983392 | 11:59923507 |
| MS4A4E | rs670139 | 11:59971794 |
| PICALM | rs10792832 | 11:85867874 |
| ABCA7 | rs4147929 | 19:1063442 |
| CD33 | rs3865444 | 19:51727961 |
| HLA-DRB5/HLA-DRB1 | rs9271192 | 6:32578529 |
| PTK2B | rs28834970 | 8:27195120 |
| SORL1 | rs11218343 | 11:121435586 |
| SLC24A4/RIN3 | rs10498633 | 14:92926951 |
| DSG2 | rs8093731 | 18:29088957 |
| INPP5D | rs35349669 | 2:234068475 |
| MEF2C | rs190982 | 5:88223419 |
| NME8 | rs2718058 | 7:37841533 |
| ZCWPW1 | rs1476679 | 7:100004445 |
| CELF1 | rs10838725 | 11:47557870 |
| FERMT2 | rs17125944 | 14:53400628 |
| CASS4 | rs7274581 | 20:55018259 |

We only included SNPs that are consistently replicated in AD GWAS studies (Lambert et al. [17] and <http://www.AlzGene.org>)

within a population of interest, compared to how often they “should” be inherited together (i.e., the difference between the observed co-occurrence of two SNPs and the expected co-occurrence of the two SNPs). LD D , the co-occurrence of events (SNPs) A_1 and B_1 (as opposed to A_2 and B_2 , respectively) is calculated as:

$$D = p(A_1 \wedge B_1) - p(A_1) * p(B_1)$$

In the simple case often seen in genetics, these events are different nucleotides (A, C, T, or G) that exist at specific positions in the genome. We would like our LD to reflect the likelihood of observing both SNPs together (e.g., if A_1 occurs then we know confidently that B_1 also occurs, and vice versa). Unfortunately, D does not provide this information. However, a second measure of LD, r^2 , based on D , is a measure of how closely related the two events (SNPs) are. A measure of $r^2 = 1.0$ means they provide the exact same information, or always co-occur. D

is converted to the Pearson correlation coefficient r by the following:

$$r = D / \sqrt{p(A_1) * p(A_2) * p(B_1) * p(B_2)}$$

In GWAS, LD is used to select which SNPs from the genome to analyze. For example, if two SNPs provide the exact same information (i.e., they always or almost always are inherited together), then only one of the SNPs is analyzed. This reduces the total number of tests (i.e., preserves statistical power) by eliminating redundant tests.

In our research, rather than using p -values from a regression to assess the relationship between a SNP (or multiple SNPs) and AD case/control status (i.e., which SNPs are correlated with AD), we calculated r^2 between each SNP in the dataset and AD case/control status. In this approach, the two events we are measuring are the co-occurrence of a SNP and AD case/control status (i.e., do case status and a particular SNP co-occur with high confidence). We accomplished this by writing our own algorithm that computes LD between each SNP and AD case/control status.

An outline of our single-SNP GWAS algorithm is given in Algorithm 1. Note that there are multiple genotype values of A_i because there are two haplotypes. The computation of Pearson's r is as described above given the computed probabilities. This approach on a set of individuals, S , and a set of SNPs, L , runs in $O(\|S\| \times \|L\|)$ time and with $O(1)$ space.

Algorithm 1 SINGLE-SNP GWAS

Input: Genotypes for a set of SNP loci, L and phenotypes for a set of individuals, S

```

1: for SNP location  $l$  in  $L$  do
2:   initialize  $Counts$ 
3:   for individual  $s \in S$  do
4:      $B_j \leftarrow$  phenotype of  $s$ 
5:     for genotype  $A_i$  at  $l$  do
6:       increment  $Counts(A_i, B_j)$ 
7:     end for
8:   end for
9:   compute  $p(A_i \wedge B_j)$  for all values of  $A_i, B_j$ 
10:  compute  $p(A_i)$  and  $p(B_j)$  for all values of  $A_i, B_j$ 
11:   $D \leftarrow p(A_1 \wedge B_1) - (p(A_1) \times p(B_1))$ 
12:  compute and report Pearson's  $r$ 
13: end for

```

Multi-SNP GWAS analysis

To extend to multi-SNP GWAS, we calculate LD between two SNPs and a phenotype. Comparing the co-occurrence of two SNPs and a trait results in eight different calculations for D of the form:

$$D_{i,j,k} = p(A_i \wedge B_j \wedge C_k) - p(A_i B_j) * p(C_k)$$

Unlike in the single-SNP case, the magnitude of D may differ between each of the eight comparisons. Because we are concerned with the possibility of a specific combination of alleles impacting the trait, we take the maximum among all eight values for D . Next we calculate r^2 using the specific combination of alleles as one event and any other combination (three possibilities, in the two SNP case) as the alternate outcome for that event:

$$d_{i,j,k} = \sqrt{p(A_i B_j) * (1 - p(A_i B_j)) * p(C_1) * p(C_2)}$$

This results in the following equation for r :

$$r = \max_{i,j,k} (D_{i,j,k} / d_{i,j,k})$$

To calculate the correlation of all SNPs with every other SNP would have exceeded our computational resources. We therefore calculated correlations between all SNPs and a subset of the SNPs that were located in genes with strongest and most consistent associations with AD (Table 1). Even this matrix was too large to fit into memory, so we computed subsections of the matrix in parallel on different machines. We found 2101 pairs of SNPs with an r^2 correlation greater than 0.04. These we plotted using R(Studio) to look for genes that had a (relatively) high correlation with other SNPs, as the relationship between two genes could provide important insights into disease processes.

Algorithm 2 MULTI-SNP PARALLELIZED GWAS

Input: Genotypes for a set of SNP loci, L and phenotypes for a set of individuals, S , and SNP IDs $START, END$

```

1: for SNP location  $l$  in  $L$  within partition  $\{START, END\}$ 
   do
2:   for SNP location  $m$  in  $L$  do
3:     initialize  $Counts$ 
4:     for individual  $s \in S$  do
5:        $C_k \leftarrow$  phenotype of  $s$ 
6:       for genotype  $A_i$  at  $l$  do
7:         for genotype  $B_j$  at  $m$  do
8:           increment  $Counts(A_i, B_j, C_k)$ 
9:         end for
10:      end for
11:    end for
12:    compute  $p(A_i \wedge B_j \wedge C_k)$  for all values of  $A_i, B_j, C_k$ 
13:    compute  $p(A_i), p(B_j)$ , and  $p(C_k)$  for all values of  $A_i, B_j, C_k$ 
14:     $D \leftarrow p(A_1 \wedge B_1 \wedge C_k) - (p(A_1 \wedge B_j) \times p(C_k))$ 
15:    compute and report Pearson's  $r$ 
16:  end for
17: end for

```

An outline of our multi-SNP parallelized GWAS algorithm is given in Algorithm 2. The required *START* and *END* allow the task to be partitioned and run in parallel. There are multiple genotype values of A_i and B_j because there are two haplotypes. Our solution on a set of individuals, S , and a set of SNPs, L , runs in $O(\|S\| \times \|L\|^2/k)$ time and with $O(\|L\|)$ space for each of k parallel runs (for our run $k = 201$ with each run allocated 32GB of RAM and a wall time of 6 h).

Results

Single-SNP GWAS analysis

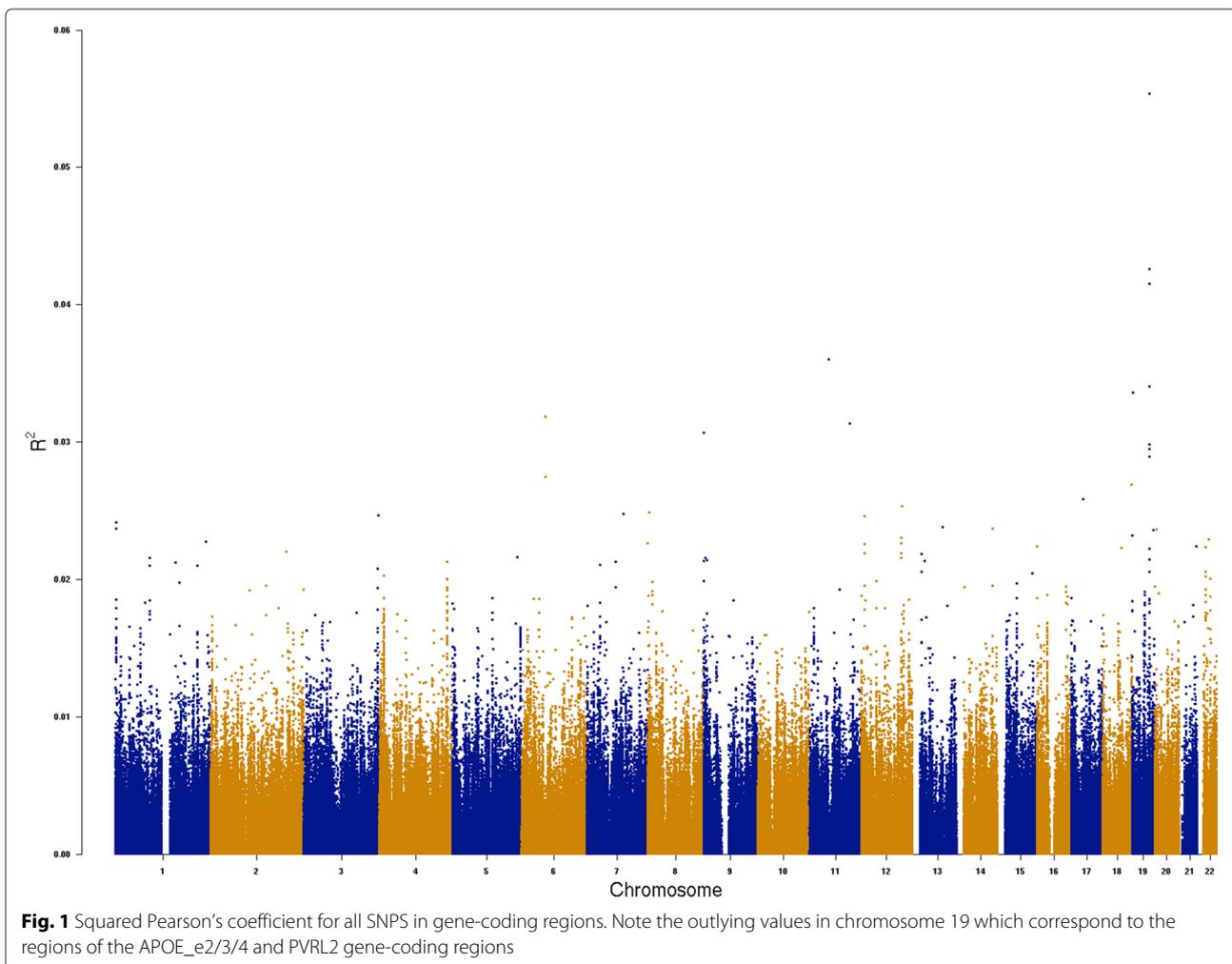
We knew in advance that our dataset would be insufficient to achieve statistical significance and present our results as a demonstration of how it could be used in a larger dataset. Although Manhattan Plots are usually calculated using p -values, we used the r^2 values to compare different SNPs (calculated using our algorithm).

For each chromosome we computed and plotted the r^2 of each SNP in all gene-coding regions (regardless of previous implication in AD). This analysis provided a comparison to traditional GWAS analyses as well as a comparison to our subsequent multi-SNP analysis (Fig. 1).

In this demonstration of our approach, we report 29 different SNPs with $r^2 \geq 0.025$ (a very conservative cutoff, which corresponds to a $p = 1.578e - 5$ (Fig. 2)) in 14 genes (Table 2). Of these 14 genes, we found 3 (APOE, TOMM40, and PVRL2) that have been previously implicated in AD.

Multi-SNP GWAS analysis

We calculated correlations between all SNPs and genic SNPs located in genes previously associated with AD (i.e. Table 1), without regard to minor allele frequency (i.e. included both rare and common SNPs). We selected 0.04 as a cutoff for the multi-SNP GWAS analysis (Fig. 3). Of



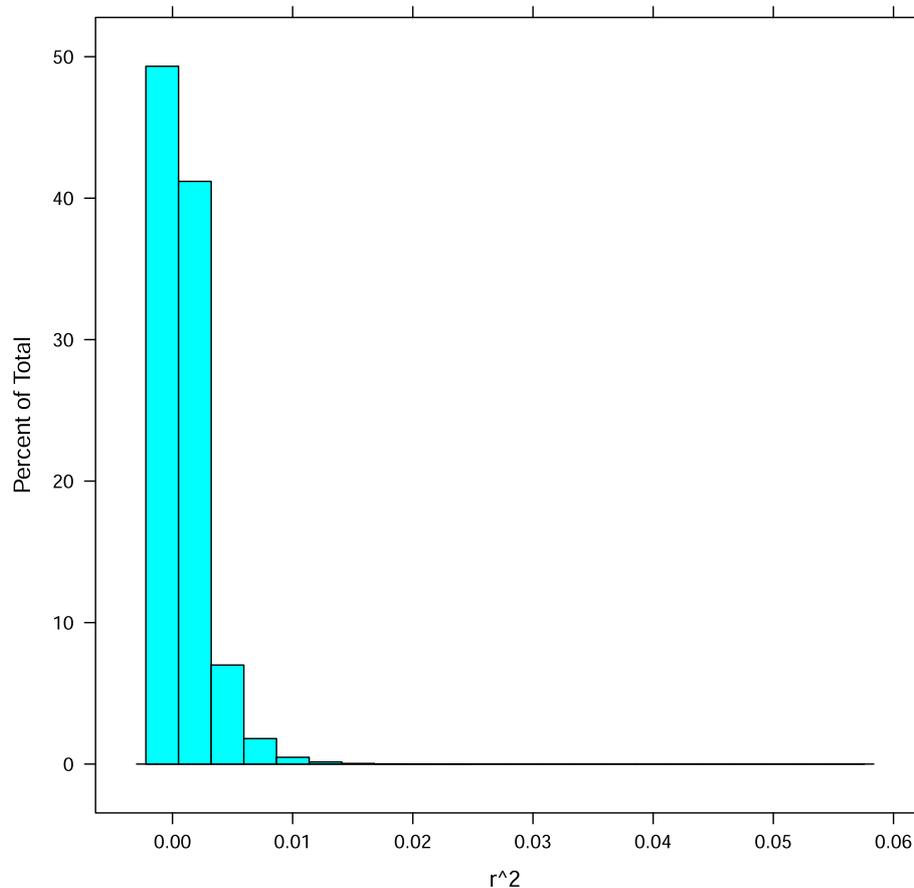


Fig. 2 Single-SNP r^2 distribution. Histogram reflecting the frequency of binned r^2 values or correlations between single SNPs and AD phenotype

4 million randomly selected SNPs from the distribution, only 3 had $r^2 \geq 0.04$ ($p = 7.501e - 7$). In practice, a Bonferroni correction could be performed by choosing different r^2 cutoffs based on the corrected alpha. We identified 192 pairs correlated with AD with an $r^2 \geq 0.04$. Additional file 1 lists the AD SNP together with the non-AD SNP and the r^2 correlation with the AD phenotype. The table is sorted by r^2 value (most to least significant). For each pair of genes, only SNP pair with the highest r^2 value was included. The final column indicates how many total significant SNP pairs were found for the given pair of genes. For example, the highest correlation had an r^2 value of 0.0730273 and was found between rs72508453 (in AD gene HLA-DRB5) and a SNP on chromosome 16 at position 6110138 (in non-AD gene RP11-509E10.1). There were 15 significant SNP pairs between these two genes. This could suggest an epistatic relationship between these two genes that is correlated with the AD phenotype.

To visualize our results, we plotted the correlation of AD SNPs (on the X axis) against all genic SNPs (on

the Y axis) in Fig. 4. Only SNP pairs with significant r^2 values are shown. Plotting the significant pairs revealed two very clear bands, and an additional three bands of interest. A vertical band indicates that an AD gene has significant epistatic interaction with other non-AD genes. The two clearest bands correspond with SNPs in HLA-DRB5/HLA-DRB1 and FERMT2, the other three bands correspond with SNPs in INPP5D, SORL1, and RIN3/SLC24A4. All but two known AD genes (MS4A6A and CD33) had correlated SNP pairs. 12 different non-AD genes had 44 or more correlated SNP pairings.

Among the 10 SNP pairs with strongest correlation, half of the pairings include a SNP from HLA-DRB5, two from INPP5D, two from FERMT2, and one from PTK2B. All but one of the SNPs from the top 10 pairs mentioned comes from a unique SNP with no prior reported AD involvement. Among the top 10, ELF4 had two SNPs with significant pairings. The first (rs6637686) was correlated with rs78623109 in FERMT2. The second (rs3788847) was correlated with rs377610860 in PTK2B.

Table 2 SNPs with $r^2 \geq 0.025$

| Gene | Chr:Pos | r^2 |
|----------------|--------------|-----------|
| KHDRBS2KHDRBS2 | 6:62728214 | 0.027466 |
| KHDRBS2 | 6:62728218 | 0.0318925 |
| CBWD1 | 9:136147 | 0.0306714 |
| RP11-347H15.1 | 11:50136388 | 0.0360252 |
| RP11-886D15.1 | 11:104457155 | 0.0313399 |
| C12orf23 | 12:107350959 | 0.0253392 |
| RP11-220C2.1 | 17:31004608 | 0.0258624 |
| NFATC1 | 18:77243594 | 0.0269173 |
| ABHD17A | 19:1879687 | 0.0336168 |
| PVRL2 | 19:45392254 | 0.0295325 |
| TOMM40 | 19:45406673 | 0.0298293 |
| APOE | 19:45410002 | 0.0415463 |
| APOE | 19:45411941 | 0.0553697 |
| APOC1 | 19:45418790 | 0.0289349 |
| APOC1 | 19:45421254 | 0.0340441 |
| APOC1 | 19:45422160 | 0.0425822 |
| RP11-265P11.1 | X:39085516 | 0.0293786 |
| RP11-265P11.1 | X:39086096 | 0.0301775 |
| RP11-265P11.1 | X:39086353 | 0.0288563 |
| RP11-265P11.1 | X:39098097 | 0.031864 |
| GS1-256O22.5 | X:142421475 | 0.0288961 |
| GS1-256O22.5 | X:142422021 | 0.0254872 |
| GS1-256O22.5 | X:142426750 | 0.0275156 |
| GS1-256O22.5 | X:142429074 | 0.0276677 |
| GS1-256O22.5 | X:142430401 | 0.026733 |
| GS1-256O22.5 | X:142439000 | 0.0268576 |
| UBL4A | X:153713451 | 0.0260031 |
| UBL4A | X:153713787 | 0.0253136 |
| UBL4A | X:153714030 | 0.0259109 |

Discussion

Although our sample size is too small to identify statistically significant correlations, this work demonstrates the utility of taking a linkage disequilibrium approach in our single- and multi-SNP GWAS. First, we identified APOE as an important AD gene in the single-SNP GWAS using our new approach. This is hardly novel as two different APOE SNPs, rs7412 and rs429358, are the most established genetic protective/risk factors for AD [8]. However, it is significant because the APOE effect is strong enough that we would expect to detect the signal in even a small dataset. In contrast, all of the other known variations that affect risk for AD are either very rare (PLD3

[11], APP [13], TREM2 [12], etc.), or have very small effect sizes [17]. Consequently, because of our sample size it is not surprising that we did not detect any other Results of the AD markers in the single-SNP GWAS.

Next, we applied our multi-SNP GWAS to identify pairs of SNPs jointly correlated with AD. In principle, epistatic interactions can exist between two genes that individually are not correlated with disease. In this study we tested the pairing of each genic SNP in known AD genes with every other SNP in the dataset. In this analysis there were five strong bands corresponding with AD genes with a large number of identified SNPxSNP pairings.

Among the 10 strongest correlations, ELF4 had two correlated SNPs (one with FERMT2 and one with HLA-DRB5). ELF4 (E74-like factor 4) is a transcription factor involved in immunity and cell cycle control [18–20]. The function of FERMT2 is unknown. However, HLA-DRB5 has a direct role in immunity [21]. From a biological standpoint, an interaction between HLA-DRB5 and ELF4 makes sense, and the immune system has a known role in AD [22].

GRIA3 (glutamate receptor, ionotropic, ampa 3) is another gene with SNPs of potential interest. A SNP in GRIA3 (rs7061304) was the second strongest correlation in our analyses (paired with rs67588672 in HLA-DRB5), and GRIA3 is among the non AD genes with the highest number of correlations. GRIA3 is a glutamate receptor. Glutamate receptors are the primary neurotransmitter receptors in human brains and GRIA3 specifically has a role in learning and memory. Furthermore, GRIA3 has been implicated in numerous disorders in the brain including bipolar disorder, mental retardation, and encephalopathy with epileptic seizures (Rasmussen encephalitis) [23, 24]. Finally, mutations in GRIA3 have been associated with cognitive impairment [25]. Although GRIA3 has not been associated with AD, glutamate receptors have been studied (including members in the same family of genes as GRIA3) for their effect on Alzheimer's disease based on the hypothesis that malfunction of glutamate receptors leads to AD-specific cell loss [26]. When considering the function of GRIA3, especially in relationship with a known AD SNP (HLA-DRB5), GRIA3 is an attractive candidate gene for further studies.

Conclusions

In summary, we developed a novel multi-SNP GWAS method and demonstrated its utility in an AD dataset. Using this approach we identified potential epistatic interactions that affect risk for AD. GRIA3, in particular, is especially intriguing and warrants followup

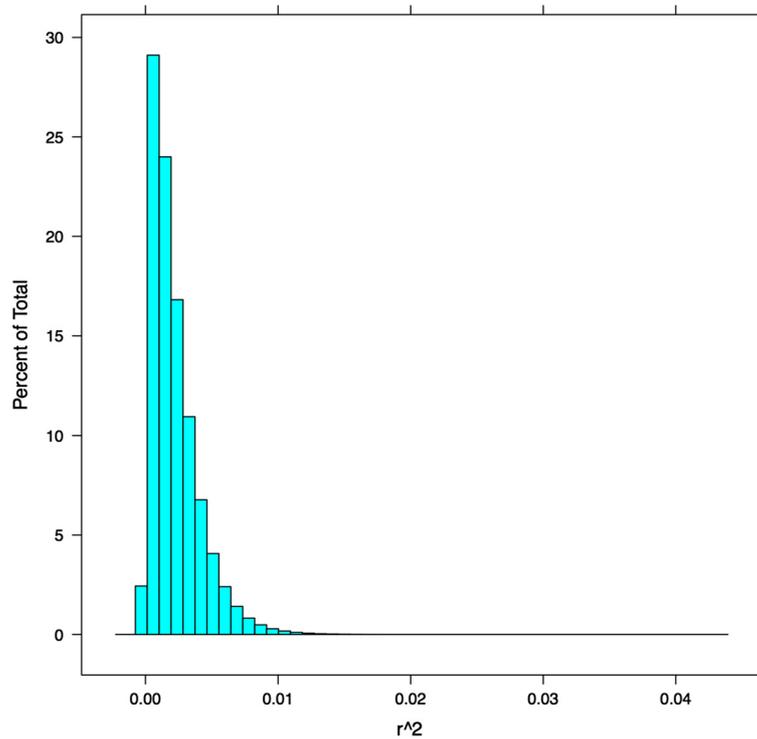


Fig. 3 Multi-SNP r^2 distribution. Histogram reflecting the frequency of binned r^2 values or correlations between pairs of SNPs and AD phenotype

studies in larger datasets. Due to the difficulty identifying epistatic interactions, relatively few interactions are known. Future work will focus on developing an appropriate dataset with experimentally validated epistatic interactions for testing new models, integrating

known biology of genes in identified interactions to identify mechanisms of synergistic functionality, and modification of the approach to more appropriately ascribe statistical significance (i.e., p -values) to identified interactions.

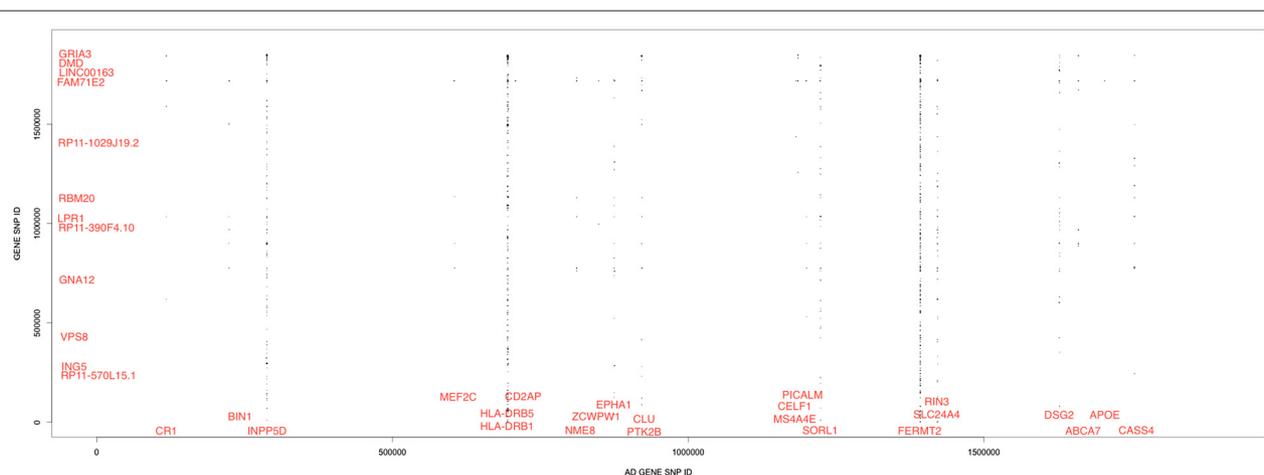


Fig. 4 Results of the Multi-SNP GWAS analysis. SNP pairs with r^2 values >0.04 were plotted. SNP ID is based on SNP order, not SNP position. For genes with several correlated SNP pairs, gene names roughly indicate corresponding SNP IDs (see Additional file 1 for more detailed information). On the X-axis, any AD gene that had correlated SNP pairings is labeled, and on the y-axis any gene that had 44 or more correlated SNP pairings is labeled. To avoid plotting all SNP positions, SNP IDs were used. SNP IDs reflect the same ordering as the associated SNP positions but are limited to include only analyzed, gene-coding SNPs

Additional file

Additional file 1: Table 3. Gene pairs with SNP pairs having $r^2 \geq 0.04$. (PDF 169 kb)

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. See http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf for a complete listing of ADNI investigators. 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.; 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Declarations

Publication of this article was funded by the Department of Biology and the College of Life Sciences at Brigham Young University. This article has been published as part of *BMC Bioinformatics* Volume 17 Supplement 7, 2016: Selected articles from the 12th Annual Biotechnology and Bioinformatics Symposium: bioinformatics.com/articles/supplements/volume-17-supplement-7. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-17-supplement-7>.

Availability of data and materials

Data is available from the Alzheimer's Disease Neuroimaging Initiative (ADNI). More information at <http://adni.loni.usc.edu/data-samples/access-data/>.

Authors' contributions

PMB, MSF, JTP formulated the study. MTWE and PGR aided with statistical analyses and biological insights. PMB, MSF, JTP, and PGR drafted the manuscript. MJC was involved in collaborative revision and direction of research. 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

Data and analyses in this manuscript were approved by the Brigham Young University Institutional Review Board.

Author details

¹Computer Science Department, Brigham Young University, 84602-6576 Provo, UT, USA. ²Department of Biology, Brigham Young University, 84602-6576 Provo, UT, USA.

Published: 25 July 2016

References

- Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch'ang L-Y, Huang W, Liu B, Shen Y, et al. The international hapmap project. *Nature*. 2003;426(6968):789–96.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9(5):356–69.
- Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY, Noureddine M, Gilbert JR, et al. Complement factor h variant increases the risk of age-related macular degeneration. *Science*. 2005;308(5720):419–21.
- Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–78.
- Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, et al. The nhgri was catalog, a curated resource of snp-trait associations. *Nucleic Acids Res*. 2014;42(D1):D1001–6.
- Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol*. 2010;8(1):34.
- Ebbert MT, Ridge PG, Kauwe JS. Bridging the gap between statistical and biological epistasis in alzheimer's disease. *BioMed Res Int*. 2015;201:121–9.
- Ridge PG, Ebbert MTW, Kauwe JSK. Genetics of alzheimer's disease. *BioMed Res Int*. 2013;2013:13. Article ID 254954. doi:10.1155/2013/254954.
- Ebbert MT, Ridge PG, Wilson AR, Sharp AR, Bailey M, Norton MC, Tschanz JT, Munger RG, Corcoran CD, Kauwe JS. Population-based analysis of alzheimer's disease risk alleles implicates genetic interactions. *Biol Psychiatry*. 2014;75(9):732–7.
- Ebbert MT, Boehme KL, Wadsworth ME, Staley LA, Mukherjee S, Crane PK, Ridge PG, Kauwe JS. Interaction between variants in *clu* and *ms4a4e* modulates alzheimer's disease risk. *Alzheimers Dement J Alzheimers Assoc*. 2015;12(2):121–9.
- Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, Harari O, Norton J, Budde J, Bertelsen S, et al. Rare coding variants in the phospholipase d3 gene confer risk for alzheimer's disease. *Nature*. 2014;505(7484):550–4.
- Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, Cruchaga C, Sassi C, Kauwe JS, Younkin S, et al. *Trem2* variants in alzheimer's disease. *N Engl J Med*. 2013;368(2):117–27.
- Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, et al. A mutation in *app* protects against alzheimer's disease and age-related cognitive decline. *Nature*. 2012;488(7409):96–9.
- Ridge PG, Mukherjee S, Crane PK, Kauwe JS, et al. Alzheimer's disease: analyzing the missing heritability. *PLoS ONE*. 2013;8(11):e79771.
- Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nat Genet*. 2011;43(5):491–8.
- Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, DeStefano AL, Bis JC, Beecham GW, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for alzheimer's disease. *Nat Genet*. 2013;45(12):1452–8.
- Miyazaki Y, Sun X, Uchida H, Zhang J, Nimer S. Mef, a novel transcription factor with an elf-1 like dna binding domain but distinct transcriptional activating properties. *Oncogene*. 1996;13(8):1721–9.
- Aryee D, Petermann R, Kos K, Henn T, Haas O, Kovar H. Cloning of a novel human elf-1-related ets transcription factor, *elfr*, its characterization and chromosomal assignment relative to *elf-1*. *Gene*. 1998;210(1):71–8.
- Yamada T, Park CS, Mamonkin M, Lacorazza HD. Transcription factor *elf4* controls the proliferation and homing of *cd8+* t cells via the *kruppel*-like factors *klf4* and *klf2*. *Nat Immunol*. 2009;10(6):618–26.
- Tieber V, Abruzzini L, Didier D, Schwartz B, Rotwein P. Complete characterization and sequence of an *hla class ii dr beta chain* cDNA from the *dr5* haplotype. *J Biol Chem*. 1986;261(6):2738–42.
- McGeer P, Akiyama H, Itagaki S, McGeer E. Immune system response in alzheimer's disease. *Can J Neurol Sci Le journal canadien des sciences neurologiques*. 1989;16(4 Suppl):516–27.

23. Gécz J, Barnett S, Liu J, Hollway G, Donnelly A, Eyre H, Eshkevari HS, Baltazar R, Grunn A, Nagaraja R, et al. Characterization of the human glutamate receptor subunit 3 gene (*gria3*), a candidate for bipolar disorder and nonspecific x-linked mental retardation. *Genomics*. 1999;62(3):356–68.
24. Rogers SW, Andrews PI, Gahring LC, Whisenand T, Cauley K, Crain B, Hughes TE, Heinemann SF, McNamara JO. Autoantibodies to glutamate receptor *glur3* in rasmussen's encephalitis. *Science*. 1994;265(5172):648–51.
25. Wu Y, Arai AC, Rumbaugh G, Srivastava AK, Turner G, Hayashi T, Suzuki E, Jiang Y, Zhang L, Rodriguez J, et al. Mutations in ionotropic ampa receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. *Proc Natl Acad Sci*. 2007;104(46):18163–8.
26. Pellegrini-Giampietro D, Bennett M, Zukin R. Ampa/kainate receptor gene expression in normal and alzheimer's disease hippocampus. *Neuroscience*. 1994;61(1):41–9.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

