

## Alzheimer's disease: analyzing the missing heritability

Perry G. Ridge<sup>a,b</sup>, Shubhabrata Mukherjee<sup>c</sup>, Paul Crane<sup>c</sup>, John S.K. Kauwe<sup>a\*</sup>, Alzheimer's

Disease Genetics Consortium<sup>\*\*</sup>

<sup>a</sup>Department of Biology, Brigham Young University, Provo, UT, USA, 84602

<sup>b</sup>ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT, USA,  
84108

<sup>c</sup>Department of Medicine, University of Washington, Seattle, WA, USA, 98122

\*To whom correspondence should be addressed

\*\*Coauthors from the Alzheimer's Disease Genetics Consortium are listed below

Perry G. Ridge: [perry.ridge@byu.edu](mailto:perry.ridge@byu.edu)

Shubhabrata Mukherjee: [smukherj@u.washington.edu](mailto:smukherj@u.washington.edu)

Paul Crane: [pcrane@uw.edu](mailto:pcrane@uw.edu)

John S.K. Kauwe: [kauwe@byu.edu](mailto:kauwe@byu.edu), 801-422-2993, 801-422-0090 (fax)

### ADGC Authors and Affiliations:

Marilyn S. Albert<sup>1</sup>, Roger L. Albin<sup>2-4</sup>, Liana G. Apostolova<sup>5</sup>, Steven E. Arnold<sup>6</sup>, Clinton T. Baldwin<sup>7</sup>, Robert Barber<sup>8</sup>, Michael M. Barmada<sup>9</sup>, Lisa L. Barnes<sup>10, 11</sup>, Thomas G. Beach<sup>12</sup>, Gary W. Beecham<sup>13, 14</sup>, Duane Beekly<sup>15</sup>, David A. Bennett<sup>10, 16</sup>, Eileen H. Bigio<sup>17</sup>, Thomas D. Bird<sup>18</sup>, Deborah Blacker<sup>19, 20</sup>, Bradley F. Boeve<sup>21</sup>, James D. Bowen<sup>22</sup>, Adam Boxer<sup>23</sup>, James R. Burke<sup>24</sup>, Joseph D. Buxbaum<sup>25, 26, 27</sup>, Nigel J. Cairns<sup>28</sup>, Laura B. Cantwell<sup>29</sup>, Chuanhai Cao<sup>30</sup>, Chris S. Carlson<sup>31</sup>, Regina M. Carney<sup>32</sup>, Minerva M. Carrasquillo<sup>33</sup>, Steven L. Carroll<sup>34</sup>, Helena C. Chui<sup>35</sup>, David G. Clark<sup>36</sup>, Jason Corneveaux<sup>37</sup>, Paul K. Crane<sup>38</sup>, David H. Cribbs<sup>39</sup>, Elizabeth A. Crocco<sup>40</sup>, Carlos Cruchaga<sup>41</sup>, Philip L. De Jager<sup>42, 43</sup>, Charles DeCarli<sup>44</sup>, Steven T. DeKosky<sup>45</sup>, F. Yesim Demirci<sup>9</sup>, Malcolm Dick<sup>46</sup>, Dennis W. Dickson<sup>33</sup>, Ranjan Duara<sup>47</sup>, Nilufer Ertekin-Taner<sup>33, 48</sup>, Denis Evans<sup>49</sup>, Kelley M. Faber<sup>50</sup>, Kenneth B. Fallon<sup>34</sup>, Martin R. Farlow<sup>51</sup>, Lindsay A Farrer<sup>7, 52, 76, 77, 83</sup>, Steven Ferris<sup>53</sup>, Tatiana M. Foroud<sup>50</sup>, Matthew P. Frosch<sup>54</sup>, Douglas R. Galasko<sup>55</sup>, Mary Ganguli<sup>56</sup>, Marla Gearing<sup>57, 58</sup>, Daniel H. Geschwind<sup>59</sup>, Bernardino Ghetti<sup>60</sup>, John R. Gilbert<sup>13, 14</sup>, Sid Gilman<sup>2</sup>, Jonathan D. Glass<sup>61</sup>, Alison M.

Goate<sup>41</sup>, Neill R. Graff-Radford<sup>33,48</sup>, Robert C. Green<sup>62</sup>, John H. Growdon<sup>63</sup>, Jonathan L. Haines<sup>64, 65</sup>, Hakon Hakonarson<sup>66</sup>, Kara L. Hamilton-Nelson<sup>13</sup>, Ronald L. Hamilton<sup>67</sup>, John Hardy<sup>68</sup>, Lindy E. Harrell<sup>36</sup>, Elizabeth Head<sup>69</sup>, Lawrence S. Honig<sup>70</sup>, Matthew J. Huentelman<sup>37</sup>, Christine M. Hulette<sup>71</sup>, Bradley T. Hyman<sup>63</sup>, Gail P. Jarvik<sup>72,73</sup>, Gregory A. Jicha<sup>74</sup>, Lee-Way Jin<sup>75</sup>, Gyungah Jun<sup>7,76,77</sup>, M. Ilyas Kamboh<sup>9,78</sup>, Anna Karydas<sup>23</sup>, John S.K. Kauwe<sup>79</sup>, Jeffrey A. Kaye<sup>80,81</sup>, Ronald Kim<sup>82</sup>, Edward H. Koo<sup>55</sup>, Neil W. Kowall<sup>83,84</sup>, Joel H. Kramer<sup>85</sup>, Patricia Kramer<sup>80,86</sup>, Walter A. Kukull<sup>87</sup>, Frank M. LaFerla<sup>88</sup>, James J. Lah<sup>61</sup>, Eric B. Larson<sup>38,89</sup>, James B. Leverenz<sup>90</sup>, Allan I. Levey<sup>61</sup>, Ge Li<sup>91</sup>, Andrew P. Lieberman<sup>92</sup>, Chiao-Feng Lin<sup>29</sup>, Oscar L. Lopez<sup>78</sup>, Kathryn L. Lunetta<sup>76</sup>, Constantine G. Lyketsos<sup>93</sup>, Wendy J. Mack<sup>94</sup>, Daniel C. Marson<sup>36</sup>, Eden R. Martin<sup>13,14</sup>, Frank Martiniuk<sup>95</sup>, Deborah C. Mash<sup>96</sup>, Eliezer Masliah<sup>55,97</sup>, Richard Mayeux<sup>70, 109, 110</sup>, Wayne C. McCormick<sup>38</sup>, Susan M. McCurry<sup>98</sup>, Andrew N. McDavid<sup>31</sup>, Ann C. McKee<sup>83,84</sup>, Marsel Mesulam<sup>99</sup>, Bruce L. Miller<sup>23</sup>, Carol A. Miller<sup>100</sup>, Joshua W. Miller<sup>75</sup>, Thomas J. Montine<sup>90</sup>, John C. Morris<sup>28, 101</sup>, Jill R. Murrell<sup>50, 60</sup>, Amanda J. Myers<sup>40</sup>, Adam C. Naj<sup>13</sup>, John M. Olichney<sup>44</sup>, Vernon S. Pankratz<sup>102</sup>, Joseph E. Parisi<sup>103,104</sup>, Margaret A. Pericak-Vance<sup>13, 14</sup>, Elaine Peskind<sup>91</sup>, Ronald C. Petersen<sup>21</sup>, Aimee Pierce<sup>39</sup>, Wayne W. Poon<sup>46</sup>, Huntington Potter<sup>30</sup>, Joseph F. Quinn<sup>80</sup>, Ashok Raj<sup>30</sup>, Murray Raskind<sup>91</sup>, Eric M. Reiman<sup>37,105-107</sup>, Barry Reisberg<sup>53,108</sup>, Christiane Reitz<sup>70,109,110</sup>, John M. Ringman<sup>5</sup>, Erik D. Roberson<sup>36</sup>, Ekaterina Rogaeva<sup>111</sup>, Howard J. Rosen<sup>23</sup>, Roger N. Rosenberg<sup>112</sup>, Mary Sano<sup>26</sup>, Andrew J. Saykin<sup>50,113</sup>, Gerard D. Schellenberg<sup>29</sup>, Julie A. Schneider<sup>10,114</sup>, Lon S. Schneider<sup>35,115</sup>, William W. Seeley<sup>23</sup>, Amanda G. Smith<sup>30</sup>, Joshua A. Sonnen<sup>90</sup>, Salvatore Spina<sup>60</sup>, Peter St George-Hyslop<sup>111,116</sup>, Robert A. Stern<sup>83</sup>, Rudolph E. Tanzi<sup>63</sup>, John Q. Trojanowski<sup>29</sup>, Juan C. Troncoso<sup>117</sup>, Debby W. Tsuang<sup>91</sup>, Otto Valladares<sup>29</sup>, Vivianna M. Van Deerlin<sup>29</sup>, Linda J. Van Eldik<sup>118</sup>, Badri N. Vardarajan<sup>7</sup>, Harry V. Vinters<sup>5,119</sup>, Jean Paul Vonsattel<sup>120</sup>, Li-San Wang<sup>29</sup>, Sandra Weintraub<sup>99</sup>, Kathleen A. Welsh-Bohmer<sup>24, 121</sup>, Jennifer Williamson<sup>70</sup>, Randall L. Woltjer<sup>122</sup>, Clinton B. Wright<sup>123</sup>, Steven G. Younkin<sup>33</sup>, Chang-En Yu<sup>38</sup>, Lei Yu<sup>10</sup>

<sup>1</sup>Department of Neurology, Johns Hopkins University, Baltimore, Maryland, <sup>2</sup>Department of Neurology, University of Michigan, Ann Arbor, Michigan, <sup>3</sup>Geriatric Research, Education and Clinical Center (GRECC), VA Ann Arbor Healthcare System (VAAHS), Ann Arbor, Michigan, <sup>4</sup>Michigan Alzheimer Disease Center, Ann Arbor, Michigan, <sup>5</sup>Department of Neurology, University of California Los Angeles, Los Angeles, California, <sup>6</sup>Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, <sup>7</sup>Department of Medicine (Genetics Program), Boston University, Boston, Massachusetts, <sup>8</sup>Department of Pharmacology and Neuroscience, University of North Texas Health Science Center, Fort Worth, Texas, <sup>9</sup>Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, <sup>10</sup>Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, <sup>11</sup>Department of Behavioral Sciences, Rush University Medical Center, Chicago, Illinois, <sup>12</sup>Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Phoenix, Arizona, <sup>13</sup>The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, <sup>14</sup>Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida, <sup>15</sup>National Alzheimer's Coordinating Center, University of Washington, Seattle, Washington, <sup>16</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, <sup>17</sup>Department of Pathology,

Northwestern University, Chicago, Illinois, <sup>18</sup>Department of Neurology, University of Washington, Seattle, Washington, <sup>19</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, <sup>20</sup>Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, <sup>21</sup>Department of Neurology, Mayo Clinic, Rochester, Minnesota, <sup>22</sup>Swedish Medical Center, Seattle, Washington, <sup>23</sup>Department of Neurology, University of California San Francisco, San Francisco, California, <sup>24</sup>Department of Medicine, Duke University, Durham, North Carolina, <sup>25</sup>Department of Neuroscience, Mount Sinai School of Medicine, New York, New York, <sup>26</sup>Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, <sup>27</sup>Departments of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, <sup>28</sup>Department of Pathology and Immunology, Washington University, St. Louis, Missouri, <sup>29</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, <sup>30</sup>USF Health Byrd Alzheimer's Institute, University of South Florida, Tampa, Florida, <sup>31</sup>Fred Hutchinson Cancer Research Center, Seattle, Washington, <sup>32</sup>Department of Psychiatry, Vanderbilt University, Nashville, Tennessee, <sup>33</sup>Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, <sup>34</sup>Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama, <sup>35</sup>Department of Neurology, University of Southern California, Los Angeles, California, <sup>36</sup>Department of Neurology, University of Alabama at Birmingham, Birmingham, Alabama, <sup>37</sup>Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Arizona, <sup>38</sup>Department of Medicine, University of Washington, Seattle, Washington, <sup>39</sup>Department of Neurology, University of California Irvine, Irvine, California, <sup>40</sup>Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, Florida, <sup>41</sup>Department of Psychiatry and Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University School of Medicine, St. Louis, Missouri, <sup>42</sup>Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology & Psychiatry, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, <sup>43</sup>Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, <sup>44</sup>Department of Neurology, University of California Davis, Sacramento, California, <sup>45</sup>University of Virginia School of Medicine, Charlottesville, Virginia, <sup>46</sup>Institute for Memory Impairments and Neurological Disorders, University of California Irvine, Irvine, California, <sup>47</sup>Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, Florida, <sup>48</sup>Department of Neurology, Mayo Clinic, Jacksonville, Florida, <sup>49</sup>Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois, <sup>50</sup>Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana, <sup>51</sup>Department of Neurology, Indiana University, Indianapolis, Indiana, <sup>52</sup>Department of Epidemiology, Boston University, Boston, Massachusetts, <sup>53</sup>Department of Psychiatry, New York University, New York, New York, <sup>54</sup>C.S. Kubik Laboratory for Neuropathology, Massachusetts General Hospital, Charlestown, Massachusetts, <sup>55</sup>Department of Neurosciences, University of California San Diego, La Jolla, California, <sup>56</sup>Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, <sup>57</sup>Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, <sup>58</sup>Emory Alzheimer's Disease Center, Emory University, Atlanta, Georgia, <sup>59</sup>Neurogenetics

Program, University of California Los Angeles, Los Angeles, California, <sup>60</sup>Department of Pathology and Laboratory Medicine, Indiana University, Indianapolis, Indiana, <sup>61</sup>Department of Neurology, Emory University, Atlanta, Georgia, <sup>62</sup>Division of Genetics, Department of Medicine and Partners Center for Personalized Genetic Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, <sup>63</sup>Department of Neurology, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, <sup>64</sup>Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, <sup>65</sup>Vanderbilt Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, <sup>66</sup>Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, <sup>67</sup>Department of Pathology (Neuropathology), University of Pittsburgh, Pittsburgh, Pennsylvania, <sup>68</sup>Institute of Neurology, University College London, Queen Square, London, <sup>69</sup>Sanders-Brown Center on Aging, Department of Molecular and Biomedical Pharmacology, University of Kentucky, Lexington, Kentucky, <sup>70</sup>Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, New York, <sup>71</sup>Department of Pathology, Duke University, Durham, North Carolina, <sup>72</sup>Department of Genome Sciences, University of Washington, Seattle, Washington, <sup>73</sup>Department of Medicine (Medical Genetics), University of Washington, Seattle, Washington, <sup>74</sup>Sanders-Brown Center on Aging, Department Neurology, University of Kentucky, Lexington, Kentucky, <sup>75</sup>Department of Pathology and Laboratory Medicine, University of California Davis, Sacramento, California, <sup>76</sup>Department of Biostatistics, Boston University, Boston, Massachusetts, <sup>77</sup>Department of Ophthalmology, Boston University, Boston, Massachusetts, <sup>78</sup>University of Pittsburgh Alzheimer's Disease Research Center, Pittsburgh, Pennsylvania, <sup>79</sup>Department of Biology, Brigham Young University, Provo, Utah, <sup>80</sup>Department of Neurology, Oregon Health & Science University, Portland, Oregon, <sup>81</sup>Department of Neurology, Portland Veterans Affairs Medical Center, Portland, Oregon, <sup>82</sup>Department of Pathology and Laboratory Medicine, University of California Irvine, Irvine, California, <sup>83</sup>Department of Neurology, Boston University, Boston, Massachusetts, <sup>84</sup>Department of Pathology, Boston University, Boston, Massachusetts, <sup>85</sup>Department of Neuropsychology, University of California San Francisco, San Francisco, California, <sup>86</sup>Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, Oregon, <sup>87</sup>Department of Epidemiology, University of Washington, Seattle, Washington, <sup>88</sup>Department of Neurobiology and Behavior, University of California Irvine, Irvine, California, <sup>89</sup>Group Health Research Institute, Group Health, Seattle, Washington, <sup>90</sup>Department of Pathology, University of Washington, Seattle, Washington, <sup>91</sup>Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, <sup>92</sup>Department of Pathology, University of Michigan, Ann Arbor, Michigan, <sup>93</sup>Department of Psychiatry, Johns Hopkins University, Baltimore, Maryland, <sup>94</sup>Department of Preventive Medicine, University of Southern California, Los Angeles, California, <sup>95</sup>Department of Medicine - Pulmonary, New York University, New York, New York, <sup>96</sup>Department of Neurology, University of Miami, Miami, Florida, <sup>97</sup>Department of Pathology, University of California San Diego, La Jolla, California, <sup>98</sup>School of Nursing Northwest Research Group on Aging, University of Washington, Seattle, Washington, <sup>99</sup>Cognitive Neurology and Alzheimer's Disease Center, Northwestern University, Chicago, Illinois, <sup>100</sup>Department of Pathology, University of Southern California, Los Angeles, California, <sup>101</sup>Department of Neurology,

Washington University, St. Louis, Missouri, <sup>102</sup>Department of Biostatistics, Mayo Clinic, Rochester, Minnesota, <sup>103</sup>Department of Anatomic Pathology, Mayo Clinic, Rochester, Minnesota, <sup>104</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, <sup>105</sup>Arizona Alzheimer's Consortium, Phoenix, Arizona, <sup>106</sup>Department of Psychiatry, University of Arizona, Phoenix, Arizona, <sup>107</sup>Banner Alzheimer's Institute, Phoenix, Arizona, <sup>108</sup>Alzheimer's Disease Center, New York University, New York, New York, <sup>109</sup>Gertrude H. Sergievsky Center, Columbia University, New York, New York, <sup>110</sup>Department of Neurology, Columbia University, New York, New York, <sup>111</sup>Tanz Centre for Research in Neurodegenerative Disease, University of Toronto, Toronto, Ontario, <sup>112</sup>Department of Neurology, University of Texas Southwestern, Dallas, Texas, <sup>113</sup>Department of Radiology and Imaging Sciences, Indiana University, Indianapolis, Indiana, <sup>114</sup>Department of Pathology (Neuropathology), Rush University Medical Center, Chicago, Illinois, <sup>115</sup>Department of Psychiatry, University of Southern California, Los Angeles, California, <sup>116</sup>Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge, <sup>117</sup>Department of Pathology, Johns Hopkins University, Baltimore, Maryland, <sup>118</sup>Sanders-Brown Center on Aging, Department of Anatomy and Neurobiology, University of Kentucky, Lexington, Kentucky, <sup>119</sup>Department of Pathology & Laboratory Medicine, University of California Los Angeles, Los Angeles, California, <sup>120</sup>Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Pathology, Columbia University, New York, New York, <sup>121</sup>Department of Psychiatry & Behavioral Sciences, Duke University, Durham, North Carolina, <sup>122</sup>Department of Pathology, Oregon Health & Science University, Portland, Oregon, <sup>123</sup>Evelyn F. McKnight Brain Institute, Department of Neurology, Miller School of Medicine, University of Miami, Miami, Florida

## **Abstract**

Alzheimer's disease (AD) is a complex disorder influenced by environmental and genetic factors. Recent work has identified 11 AD markers in 10 loci. We used Genome-wide Complex Trait Analysis to analyze >2 million SNPs for 10,922 individuals from the Alzheimer's Disease Genetics Consortium to assess the phenotypic variance explained first by known late-onset AD loci, and then by all SNPs in the Alzheimer's Disease Genetics Consortium dataset. In all, 33% of total phenotypic variance is explained by all common SNPs. *APOE* alone explained 6% and other known markers 2%, meaning more than 25% of phenotypic variance remains unexplained by known markers, but is tagged by common SNPs included on genotyping arrays or imputed with HapMap genotypes. Novel AD markers that explain large amounts of phenotypic variance are likely to be rare and unidentifiable using genome-wide association studies. Based on our findings and the current direction of human genetics research, we suggest specific study designs for future studies to identify the remaining heritability of Alzheimer's disease.

## **Key Words**

Alzheimer's disease genetics; *APOE*; heritability; phenotypic variance; Alzheimer's disease markers; next-generation sequencing; rare variants; study design

## 1. Introduction

Alzheimer's disease (AD) is the most common form of dementia. Worldwide estimates of prevalence vary, with estimates of 24 to 35 million people affected [1-3]. Combined with an aging population, prevalence is expected to increase to 1 in 85 people by 2050 [2].

AD is a heterogeneous disease caused by a combination of environmental and genetic factors. The most important risk factor for Alzheimer's disease is age [1,4].

Environmental risk factors include hypertension, estrogen supplements [5], smoking [6,7], stroke, heart disease, depression, arthritis, and diabetes [8]. In addition, certain lifestyle choices appear to decrease the risk of AD: exercise [9], intellectual stimulation [10], and maintaining a Mediterranean diet (including fish)[11,12].

The genetics of AD are complex. Several genes are known to harbor either causative or risk variants for AD. There are two primary types of AD as defined by age. The first is early-onset, or familial AD, and the second type is late-onset AD (LOAD), sometimes termed sporadic AD. Three genes, *APP* [13], *PSEN1* [14], and *PSEN2* [15] are known to harbor many highly penetrant, autosomal dominantly-inherited variants, which lead to early-onset AD but account for only a small fraction of total AD cases.

LOAD accounts for 99% of AD cases and is caused by a more complex underlying genetic architecture. Genome-wide association studies (GWAS) have identified 10 different loci associated with AD (Table 1). Recent applications of next-generation sequencing (NGS) have suggested rare variants play important role and have large effects

in the etiology of AD [16-18]. Identifying additional variants will provide information that is integral to the development, evaluation and application of effective therapeutic strategies for AD. Lee et al. [19] used 3,333 cases and 3,924 controls, including 2,699 population-based controls to estimate that common genetic variants account for 24% of variance in AD. They also estimated the contribution of APOE using several proxy SNPs, with varying degrees of LD, with the APOE  $\epsilon$ 4 allele to estimate the APOE effect at approximately 4%. Here we evaluate the variance in AD status explained by common SNPs and along with all recently identified AD genes, including direct genotyping of the APOE  $\epsilon$ 2 and  $\epsilon$ 4 alleles, in 5,708 AD cases and 5,214 clinically ascertained controls. We also suggest strategies for identifying the remaining AD genes.



## 2. Methods

### 2.1 Dataset

We used the Alzheimer's Disease Genetics Consortium (ADGC) dataset described in Naj et al. [20] for our analyses. Samples were genotyped using Affymetrix and Illumina SNP chips. Quality control of the imputed data was performed as described by Naj et al. 2011 [20]. Briefly, markers with a minor allele frequency of less than 1% and deviation from HWE where  $P < 10^{-6}$  were removed. To have a common set of SNPs across all samples, imputation to HapMap phase 2 (release 22)[21] was performed using MaCH [22] and strand ambiguous SNPs were removed, resulting in a rectangular dataset with 2,042,114 SNPs. Only SNPs imputed with  $R^2 \geq 0.50$  were included in the dataset. We added an additional two SNPs, rs7412 and rs429358, corresponding to *APOE*  $\epsilon 2$  and  $\epsilon 4$ , respectively.

We used a compiled dataset of directly genotyped SNPs common to all 15 studies to assess cryptic relatedness and calculate principal components to account for population-specific variations in allele distribution. We excluded strand ambiguous SNPs, resulting in a rectangular dataset with 21,880 directly observed (not imputed) SNPs in common across all the studies. We filtered SNPs with pairwise LD ( $r^2$ )  $< 0.20$ , resulting in a dataset with 17,054 SNPs. We used both PLINK [23] and KING-ROBUST [24] for relatedness analysis. KING-ROBUST provided unbiased kinship coefficient estimates for related individuals in our dataset. We excluded up to 3<sup>rd</sup> degree relatives (kinship  $\geq 0.0442$ ) for a final dataset containing 19,692 individuals.

Of the 19,692 individuals in the original dataset we analyzed a subset of 10,922 individuals who had complete data for the 11 markers listed in Table 1, AD case-control status, age, sex, and 10 principal components from the population stratification analysis (missingness rates for each of the covariates and case-control status are reported in Table S1). Basic demographic information for the 10,922 individuals in the subset of the dataset used in this study is presented in Table 2. We collected chromosome length and number of genes per chromosome from the Vega database [25].

## ***2.2 Genetic Analyses***

We used Genome-wide Complex Trait Analysis (GCTA)[26], a tool that implements the methods described in Yang et al. [27], Lee et al. [28], and Yang et al. [29] to estimate the phenotypic variance explained by known AD genes and tagged by SNPs on the SNP arrays. Briefly, GCTA uses a mixed linear model and treats the effects of SNPs as random effects, effectively testing all the SNPs together for effect (in contrast to GWAS, which considers each SNP individually). We used age, sex, and 10 principal components as covariates. For the analyses in which we examined unexplained phenotypic variance, we also controlled for the 11 known AD markers (Table 1). The 11 known AD markers are the AlzGene.org top hits and are the markers with replicable evidence for association with AD. Each of these markers is present in our dataset except rs9349407 in CD2AP. As proxy we used rs9296559, which is in very high LD with rs9349407 ( $r^2=1$ ). We specified a population prevalence of LOAD at 0.13 [30].

### ***2.3 Ethics Statement***

All study procedures were approved by the Institutional Review Boards of Brigham Young University and the University of Washington.

### **3. Results**

We estimated the variance in AD case-control status focusing first on the 11 known AD markers (Table 1). Together these markers account for 7.8% (standard error 0.03) of the phenotypic variance (Table 3). Next, we estimated the explained phenotypic variance for each chromosome (Figure 1). Chromosome 19 accounts for the highest proportion of phenotypic variance.

In all, the 2,042,116 SNPs in the HapMap imputed ADGC dataset explain 33.1% of phenotypic variance (genetic variance of 0.0711, standard error 0.0072). The APOE  $\epsilon$ 2 and  $\epsilon$ 4 alleles account for 5.9% (standard error 0.03) of the phenotypic variance (Table 3). The other 9 known high frequency SNPs identified in GWAS explain an additional 1.9% (standard error 0.03)(Table 3). After controlling for these 11 markers, an additional 25.3% of the total phenotypic variance (genetic variance of 0.046, standard error 0.006) is explained with as-yet unidentified variants (Table 3). The remaining phenotypic variance explained by each chromosome after controlling for the 11 known markers is shown in Figure 2. SNPs on chromosomes 1, 4, 5 and 17 account for the largest percentage of remaining unexplained phenotypic variance compared to other

chromosomes, each accounting for more than 2% (Figure 2). Chromosomes 9, 14, and 21 account for the least ( $<0.0001\%$  each); however, there is unexplained phenotypic variance on all the autosomes. There is no relationship between explained variance and chromosome length (p-value = 0.8), or number of genes per chromosome (p-value = 0.7).

#### **4. Discussion**

A clear understanding of the genetic architecture of Alzheimer's disease provides the foundation of information needed to cure this terrible disease. While many large GWAS for AD have been performed and several replicable loci have been identified (as referenced in Table 1), relatively little phenotypic variance is explained by these variants. Our estimates of phenotype variance explained by common genetic variants and by the APOE locus are higher than those of Lee et al. [19]. We estimated total phenotypic variance explained by common SNPs to be ~33%. In contrast Lee et al. [19] estimated ~24%. In our study we used genotyped and HapMap imputed SNPs, whereas Lee et al. [19] used only directly genotyped SNPs. Inclusion of imputed SNPs improved heritability estimates and suggests that imputed SNPs should be included in such studies. In addition to using imputed variants, our dataset was larger and our controls were clinically ascertained. Differences in the estimates for APOE (~6% in this study compared to ~4% in Lee et al. [19]) could be due to these same characteristics as well as the direct genotyping of the APOE  $\epsilon 2$  and  $\epsilon 4$  alleles in our samples as opposed to the use of proxy markers. Regardless, both studies provide evidence that a considerable amount of variance in AD is explained by yet unidentified genetic variation. Together these results

show that there is clearly much work to be done if we are to solve the genetic architecture of AD. GWAS with sample sizes performed to date are able to identify common variants with moderate to small effect sizes. Results of GWAS in AD and other conditions suggest there may be a large number of such variants and that additional variants of this type can be identified by increasing sample sizes. However, additional loci detected with the GWAS strategy will likely have effects either similar to or smaller than SNPs already identified. The range of SNPs identifiable by current GWAS [20,31-36] is marked on Figure 3 by the large box bordered by dots (the GWAS search space), with recent GWAS hits inside the labeled oval. The GWAS being conducted by the International Genomics of Alzheimer's Project represents a substantial increase in sample size and will undoubtedly identify additional common loci with small effects on AD risk. Nevertheless, it is unlikely that many common variants of even modest effect size remain to be identified.

There are still many AD variants that remain to be identified, however, and these variants exist on every autosome (Figure 2). Variants with large effects are almost certainly present in very low frequencies or they would have been identified in GWAS. While such variants are unlikely to be detected using traditional GWAS due to limitations of  $r^2$  based "tagging" for alleles with different frequencies [37] the current analysis allows for high  $D'$  values between common alleles and rare variants of large effect to contribute to the explained variance. These rare variants of large effect appear in the smaller box bordered with dashes in Figure 3. To date, identified alleles of this type have clear functional effects and large effect sizes compared to associated alleles from GWAS. Detecting rare

variants of large effect requires different experimental designs than GWAS such as sequencing causal loci. Exome chip array studies target known variation in coding regions, even those of very low frequency; this may prove a promising and economical approach. However, accurately genotyping variants of less than 1-2% using these arrays is quite challenging, and for variants that are present below these frequencies other approaches are required.

Two seemingly contradictory hypotheses exist about the architecture of complex disease: the common disease/common variant hypothesis and the multiple rare variant hypothesis. In the first, many common variants of small effect size collectively explain disease risk, while in the second, rare variants, some with large effect and high penetrance, explain disease risk. However, as suggested by Singleton et al. [38] these two hypotheses are not mutually exclusive and the genetics of complex diseases are likely a hybrid of the two. Singleton et al. [38] suggests that both common and rare variants that increase or decrease disease risk are likely to be found in the same loci and coined the phrase “pleomorphic risk loci”. To date, AD genetics research has largely focused on common variants that influence disease risk, likely due to technological and financial constraints. However, the advent of next generation sequencing (NGS) and falling costs of this technology have made it possible to expand AD research to include searching for rare variants. Recently, this technology was used to identify a functional variant that protects against Alzheimer’s disease in the amyloid precursor protein (*APP*)[17]. Additionally, two groups recently used NGS to identify additional, likely functional, variants associated with AD in the triggering receptor expressed on myeloid 2 (*TREM2*) gene

[16,18]. The *TREM2* variant is present in about 1% of the general population and has a high odds ratio (2.9 to 5.1 depending on the dataset). Likewise, the *APP* variant is extremely rare (frequency of 0.00038), but confers a large protective effect on carriers. Larger scale applications of this technology and careful study design are likely to identify additional variants and further explain the remaining phenotypic variance in AD.

Family-based studies are also an effective application of NGS. These studies require carefully ascertained families and accurate pedigree data and can be used to identify high effect, low frequency variants (located in the box with longer dashes in Figure 3).

Family-based studies are especially powerful because large effect, low frequency disease-causing (or disease-modifying) sequence features, some of which may be unique to a single family, are likely to segregate, at least partially, with disease status. These approaches have not yet been extensively applied in AD research. Nevertheless, family-based studies utilizing large-scale genome or exome sequencing have recently been used to identify disease-causing variants in several Mendelian [39-41] and complex disorders [42,43].

It is also possible that gene-gene interactions account for much of the unexplained variance in AD status [44]. These interactions are widespread and common [45,46] and approaches to understand the effects of epistatic interactions exist and continue to mature [44,47]. Several interesting candidate interactions have been identified and Ebbert et al. 2013 (accepted) recently demonstrated that allowing interactions improves the diagnostic utility of the known AD markers. Unfortunately, the complexity of this problem and the

extremely large samples sizes required to perform agnostic screens for gene-gene interactions make it very difficult to conduct effective screens for these effects.

AD is a highly complex disease with substantial genetic and environmental components. Our results suggest that genetic variance accounts for ~30% of phenotypic variance, but over 75% of this phenotypic variance remains unexplained by currently identified AD genes. Future AD genetics research must leverage larger samples and novel technologies such as NGS to identify rare, high penetrant variation and gene-gene interactions that are likely to explain the remaining genetic and phenotypic variance in AD.

Genetic research in AD has followed roughly the same model as the study of other complex diseases; largely focusing on the identification of common variants of modest effect using association studies. Scientist in many disease fields have successfully identified numerous associated variants (this is a small representative sample [48-57]). The transition from a focus on common variants to a focus on the identification of low frequency variants is now underway. These rare, functionally relevant markers are often more easily characterized than common variants of small effect. This will lead to strong and testable hypotheses for the development of therapeutics, thus accelerating the progress toward effective prevention and treatment.

## **5. Acknowledgments**

Also, we would like to thank Mark T.W. Ebbert for his insightful comments on the manuscript.



### ***5.1 Alzheimer's Disease Genetics Consortium***

Biological samples and associated phenotypic data used in primary data analysis were stored at the Principal Investigator's institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD), at the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania, and the NIA Alzheimer's Disease Genetics Consortium Data Storage Site at the University of Pennsylvania.

Contributors to the genetic analysis data included Principal Investigators on projects that were individually funded by NIA, other NIH institutes, or private entities.

The National Institutes of Health, National Institute on Aging (NIH-NIA) supported this work through the following grants: ADGC, U01 AG032984, RC2 AG036528; NACC, U01 AG016976; NCRAD, U24 AG021886; NIA LOAD, U24 AG026395, U24 AG026390; Banner Sun Health Research Institute P30 AG019610; Boston University, P30 AG013846, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG017173, R01 AG025259, R01AG33193; Columbia University, P50 AG008702, R37 AG015473; Duke University, P30 AG028377, AG05128; Emory University, AG025688; Group Health Research Institute, UO1 AG06781, UO1 HG004610; Indiana University, P30 AG10133; Johns Hopkins University, P50 AG005146, R01 AG020688; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG016574; Mount Sinai School of Medicine, P50 AG005138, P01 AG002219; New York University, P30 AG08051, MO1RR00096, UL1 RR029893, 5R01AG012101, 5R01AG022374, 5R01AG013616, 1RC2AG036502, 1R01AG035137; Northwestern University, P30 AG013854; Oregon

Health & Science University, P30 AG008017, R01 AG026916; Rush University, P30 AG010161, R01 AG019085, R01 AG15819, R01 AG17917, R01 AG30146; TGen, R01 NS059873; University of Alabama at Birmingham, P50 AG016582, UL1RR02777; University of Arizona, R01 AG031581; University of California, Davis, P30 AG010129; University of California, Irvine, P50 AG016573, P50, P50 AG016575, P50 AG016576, P50 AG016577; University of California, Los Angeles, P50 AG016570; University of California, San Diego, P50 AG005131; University of California, San Francisco, P50 AG023501, P01 AG019724; University of Kentucky, P30 AG028383, AG05144; University of Michigan, P50 AG008671; University of Pennsylvania, P30 AG010124; University of Pittsburgh, P50 AG005133, AG030653; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG012300; University of Miami, R01 AG027944, AG010491, AG027944, AG021547, AG019757; University of Washington, P50 AG005136; Vanderbilt University, R01 AG019085; and Washington University, P50 AG005681, P01 AG03991. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is funded by NINDS grant # NS39764, NIMH MH60451 and by Glaxo Smith Kline. Genotyping of the TGEN2 cohort was supported by Kronos Science. The TGen series was also funded by NIA grant AG034504 to AJM, The Banner Alzheimer's Foundation, The Johnnie B. Byrd Sr. Alzheimer's Institute, the Medical Research Council, and the state of Arizona and also includes samples from the following sites: Newcastle Brain Tissue Resource (funding via the Medical Research Council, local NHS trusts and Newcastle University), MRC London Brain Bank for Neurodegenerative Diseases (funding via the Medical Research Council), South West Dementia Brain Bank (funding via numerous sources including the Higher Education

Funding Council for England (HEFCE), Alzheimer's Research Trust (ART), BRACE as well as North Bristol NHS Trust Research and Innovation Department and DeNDRoN), The Netherlands Brain Bank (funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stichting Alzheimer Onderzoek), Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona. ADNI Funding for ADNI is through the Northern California Institute for Research and Education by grants from Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., Alzheimer's Association, Alzheimer's Drug Discovery Foundation, the Dana Foundation, and by the National Institute of Biomedical Imaging and Bioengineering and NIA grants U01 AG024904, RC2 AG036535, K01 AG030514. We thank Drs. D. Stephen Snyder and Marilyn Miller from NIA who are *ex-officio* ADGC members. Support was also from the Alzheimer's Association (LAF, IIRG-08-89720; MP-V, IIRG-05-14147) and the US Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program. P.S.G.-H. is supported by Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institute of Health Research.

## ***5.2 Disclosure Statement***

The authors declare no conflicts of interest.

## References

1. Querfurth HW, LaFerla FM (2010) Alzheimer's disease. *N Engl J Med* 362: 329-344.
2. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM (2007) Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement* 3: 186-191.
3. Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, et al. (2011) Alzheimer's disease. *Lancet*.
4. Herrup K (2010) Reimagining Alzheimer's disease--an age-based hypothesis. *J Neurosci* 30: 16755-16762.
5. Patterson C, Feightner JW, Garcia A, Hsiung GY, MacKnight C, et al. (2008) Diagnosis and treatment of dementia: 1. Risk assessment and primary prevention of Alzheimer disease. *CMAJ* 178: 548-556.
6. Cataldo JK, Prochaska JJ, Glantz SA (2010) Cigarette smoking is a risk factor for Alzheimer's Disease: an analysis controlling for tobacco industry affiliation. *J Alzheimers Dis* 19: 465-480.
7. Almeida OP, Hulse GK, Lawrence D, Flicker L (2002) Smoking as a risk factor for Alzheimer's disease: contrasting evidence from a systematic review of case-control and cohort studies. *Addiction* 97: 15-28.
8. Lindsay J, Laurin D, Verreault R, Hebert R, Helliwell B, et al. (2002) Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging. *Am J Epidemiol* 156: 445-453.
9. Podewils LJ, Guallar E, Kuller LH, Fried LP, Lopez OL, et al. (2005) Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. *Am J Epidemiol* 161: 639-651.
10. Wang HX, Karp A, Winblad B, Fratiglioni L (2002) Late-life engagement in social and leisure activities is associated with a decreased risk of dementia: a longitudinal study from the Kungsholmen project. *Am J Epidemiol* 155: 1081-1087.
11. Scarmeas N, Stern Y, Mayeux R, Luchsinger JA (2006) Mediterranean diet, Alzheimer disease, and vascular mediation. *Arch Neurol* 63: 1709-1717.
12. Patterson C, Feightner J, Garcia A, MacKnight C (2007) General risk factors for dementia: a systematic evidence review. *Alzheimers Dement* 3: 341-347.
13. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, et al. (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* 349: 704-706.
14. Sherrington R, Rogaeve EI, Liang Y, Rogaeve EA, Levesque G, et al. (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* 375: 754-760.

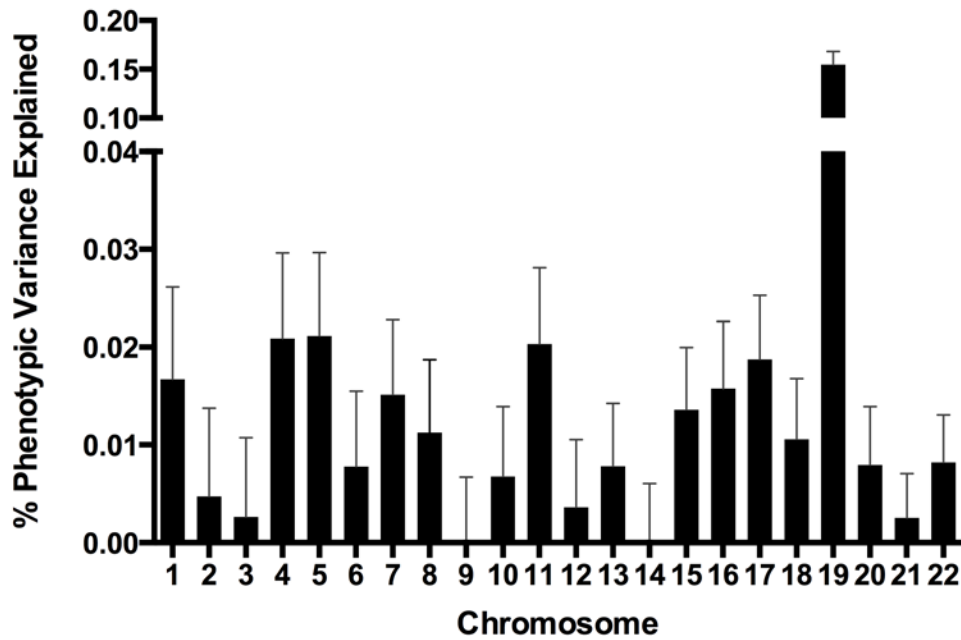
15. Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, et al. (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* 269: 973-977.
16. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, et al. (2012) TREM2 Variants in Alzheimer's Disease. *The New England journal of medicine*.
17. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, et al. (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488: 96-99.
18. Jonsson T, Stefansson H, Ph DS, Jonsdottir I, Jonsson PV, et al. (2012) Variant of TREM2 Associated with the Risk of Alzheimer's Disease. *The New England journal of medicine*.
19. Lee SH, Harold D, Nyholt DR, Goddard ME, Zondervan KT, et al. (2013) Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Human molecular genetics* 22: 832-841.
20. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, et al. (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 43: 436-441.
21. (2003) The International HapMap Project. *Nature* 426: 789-796.
22. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology* 34: 816-834.
23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575.
24. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, et al. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* 26: 2867-2873.
25. Wilming LG, Gilbert JG, Howe K, Trevanion S, Hubbard T, et al. (2008) The vertebrate genome annotation (Vega) database. *Nucleic acids research* 36: D753-760.
26. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* 88: 76-82.
27. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* 42: 565-569.
28. Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating missing heritability for disease from genome-wide association studies. *American journal of human genetics* 88: 294-305.
29. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, et al. (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics* 43: 519-525.
30. Association As (2012) Alzheimer's Association Annual Report: Alzheimer's disease Facts and Figures.

31. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, et al. (2011) Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 43: 429-435.
32. Biffi A, Anderson CD, Desikan RS, Sabuncu M, Cortellini L, et al. (2010) Genetic variation and neuroimaging measures in Alzheimer disease. *Archives of neurology* 67: 677-685.
33. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, et al. (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 41: 1088-1093.
34. Hu X, Pickering E, Liu YC, Hall S, Fournier H, et al. (2011) Meta-analysis for genome-wide association study identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease. *PLoS One* 6: e16616.
35. Carrasquillo MM, Belbin O, Hunter TA, Ma L, Bisceglia GD, et al. (2010) Replication of CLU, CR1, and PICALM associations with alzheimer disease. *Arch Neurol* 67: 961-964.
36. Corneveaux JJ, Myers AJ, Allen AN, Pruzin JJ, Ramirez M, et al. (2010) Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *Hum Mol Genet* 19: 3295-3301.
37. Wray NR (2005) Allele frequencies and the  $r^2$  measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin research and human genetics : the official journal of the International Society for Twin Studies* 8: 87-94.
38. Singleton A, Hardy J (2011) A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. *Human molecular genetics* 20: R158-162.
39. Rope AF, Wang K, Evjenth R, Xing J, Johnston JJ, et al. (2011) Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *American journal of human genetics* 89: 28-43.
40. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19096-19101.
41. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, et al. (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328: 636-639.
42. Krebs CE, Paisan-Ruiz C (2012) The use of next-generation sequencing in movement disorders. *Frontiers in genetics* 3: 75.
43. Kilpinen H, Barrett JC (2013) How next-generation sequencing is transforming complex disease genetics. *Trends in genetics : TIG* 29: 23-30.
44. Turton JC, Bullock J, Medway C, Shi H, Brown K, et al. (2011) Investigating statistical epistasis in complex disorders. *Journal of Alzheimer's disease : JAD* 25: 635-644.
45. Mackay TF (2004) The genetic architecture of quantitative traits: lessons from *Drosophila*. *Current opinion in genetics & development* 14: 253-257.
46. Shao H, Burrage LC, Sinasac DS, Hill AE, Ernest SR, et al. (2008) Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis.

- Proceedings of the National Academy of Sciences of the United States of America 105: 19910-19914.
47. Combarros O, Cortina-Borja M, Smith AD, Lehmann DJ (2009) Epistasis in sporadic Alzheimer's disease. *Neurobiology of aging* 30: 1333-1349.
  48. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, et al. (2011) Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* 378: 1006-1014.
  49. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, et al. (2011) Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature genetics* 43: 1005-1011.
  50. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, et al. (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491: 119-124.
  51. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, et al. (2012) Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS genetics* 8: e1002548.
  52. Wen W, Cho YS, Zheng W, Dorajoo R, Kato N, et al. (2012) Meta-analysis identifies common variants associated with body mass index in east Asians. *Nature genetics* 44: 307-311.
  53. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, et al. (2011) A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nature genetics* 43: 1210-1214.
  54. (2011) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature genetics* 43: 977-983.
  55. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478: 103-109.
  56. Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoeur C, et al. (2012) A genome-wide association meta-analysis identifies new childhood obesity loci. *Nature genetics* 44: 526-531.
  57. Jenuth JP (2000) The NCBI. Publicly available tools and resources on the Web. *Methods in molecular biology* 132: 301-312.
  58. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261: 921-923.
  59. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, et al. (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* 7: 180-184.
  60. Lambert JC, Heath S, Even G, Campion D, Sleegers K, et al. (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 41: 1094-1099.

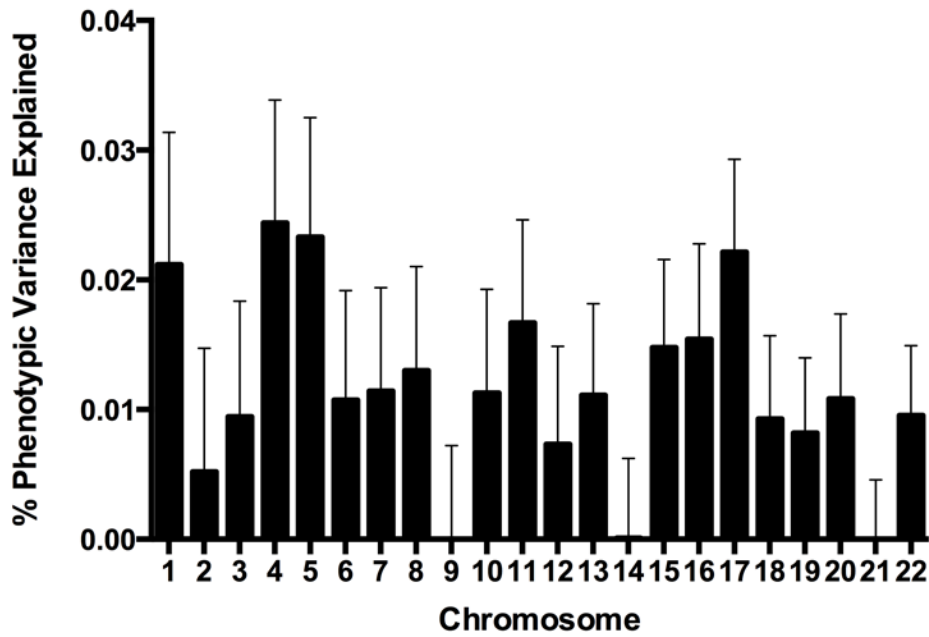
## Figure Legends

**Figure 1.** Unexplained Alzheimer's disease variance, by chromosome. In this figure we show phenotypic variance, by chromosome, explained by all SNPs. Error bars correspond to standard error.

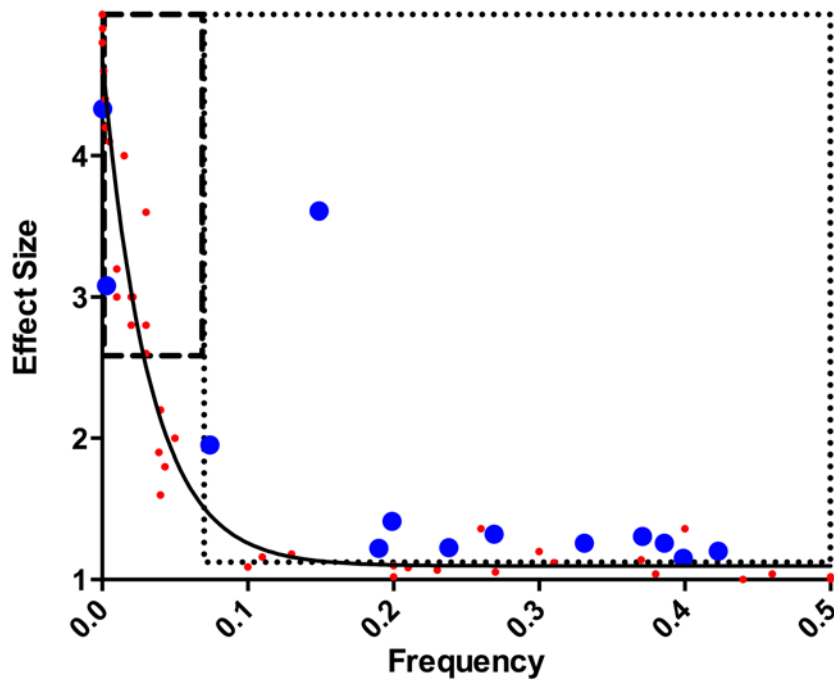




**Figure 2.** Unexplained Alzheimer’s disease variance, by chromosome, excluding known Alzheimer’s disease markers. This figure is the same as Figure 1 except we have removed variance explained by known Alzheimer’s disease markers. Error bars correspond to standard error.



**Figure 3.** Variant search space. Real and hypothetical variants are graphed by effect size (y-axis) and population frequency (x-axis). Known Alzheimer's disease SNPs are blue circles and hypothetical SNPs are red squares. The large box on the right outlined with dots, is the GWAS search space and the smaller box on the left, outlined with dashes, is the next-generation sequencing search space. SNPs identified in recent GWAS are located in the labeled oval inside the GWAS search space. Known Alzheimer's disease SNPs are those found in Table 1 as well as APP and TREM2, which are both labeled on the graph.



**Table 1.** Late-onset Alzheimer's disease associated genes/variants.

<b>Variant</b>	<b>Gene</b>	<b>Abbreviation</b>	<b>Odds Ratio</b>
rs429358	Apolipoprotein E ( $\epsilon$ 4 allele)[58]	APOE	3.685
rs7412	Apolipoprotein E ( $\epsilon$ 2 allele)[59]	APOE	0.621
rs744373	Bridging Integrator 1[32]	BIN1	1.166
rs11136000	Clusterin[33,60]	CLU	0.879
rs3764650	ATP-binding cassette, sub-family A (ABC1), member 7[31]	ABCA7	1.229
rs3818361	Complement component (3b/4b) receptor 1 (Knops blood group)[60]	CR1	1.174
rs3851179	Phosphatidylinositol binding clathrin assembly protein[33]	PICALM	0.879
rs610932	Membrane-spanning 4-domains, subfamily A, member 6A[31]	MS4A6A	0.904
rs3865444	CD33 molecule[20,31]	CD33	0.893
rs670139	Membrane-spanning 4-domains, subfamily A, member 4E[31]	MS4A4E	1.079
rs9296559	CD2-associated protein[20,31]	CD2AP	1.117

The dbSNP identification number, gene name, gene abbreviation, and odds ratio for each of the top variants from the Alzgene.org meta-analyses. The SNP in CD2AP, rs9349407, is not present in this sample, so we used rs9296559 instead as a proxy. These two SNPs are close together (1,108 base pairs apart) and in very high LD ( $r^2=1$ ).

**Table 2.** Demographic information for individuals in the analysis dataset.

	<b>Count</b>	<b>Cases / Controls</b>	<b>Age (Cases / Controls)</b>
<b>Male</b>	4489	2378/2111	74.8 (73.6 / 76.1)
<b>Female</b>	6433	3330/3103	75.3 (74.9 / 75.7)
<b>Total</b>	10922	5708/5214	75.1 (74.3 / 75.9)

We report total individuals, sex, case-control status, and average age for the 10,922 individuals analyzed in this report.

**Table 3.** Summary of genetic and phenotypic variance measurements.

<b>SNP Set</b>	<b>Genetic Variance (Standard Error)</b>	<b>% Phenotypic Variance (Standard Error)</b>
All SNPs	0.071 (0.0072)	33.12% (0.033)
APOE $\epsilon 2/\epsilon 4$	0.020 (0.0066)	5.92% (0.033)
All known AD markers (Table 1)	0.025 (0.0066)	7.78% (0.033)
All SNPs excluding known AD markers	0.046 (0.0060)	25.34% (0.033)

In this table we summarize our results, showing genetic and percent phenotypic variance for four different subsets of SNPs: all SNPs in the ADGC dataset, the two APOE alleles ( $\epsilon 2$  and  $\epsilon 4$ ), all known AD markers (as listed in Table 1), and all SNPs excluding those in Table 1.

**Supplementary Table 1. Missingness rates for covariates and case-control status**

<b>Covariates/Case-control Status</b>	<b># of individuals missing the given covariate (% missing)</b>
Age	1491 (7.57%)
Sex	0 (0%)
Case-control Status	0 (0%)
APOE (rs429358/rs7412)	579 (2.94%)
10 Principal Components	0 (0%)
BIN1 (rs744373)	2872 (14.58%)
CLU (rs11136000)	3009 (15.28%)
PICALM (rs3851179)	211 (1.07%)
MS4A6A (rs610932)	30 (0.15%)
CD33 (rs3865444)	392 (1.99%)
MS4A4E (rs670139)	192 (0.98%)
CD2AP (rs9296559)	2374 (12.06%)
CR1 (rs3818361)	54 (0.27%)
ABCA7 (rs3764650)	2964 (15.05%)

The Alzheimer's Disease Genetics Consortium dataset consists of 19,692 total individuals. We removed any individuals missing any of the covariates (listed here) or case-control status (included in this table).