Supplementary Online Content


eAppendix 1. List of Investigators
eAppendix 2. Additional Information
eReferences
eFigure 1. Conditional quantile-quantile (Q-Q) plots after removing all SNPs located in the MHC (location 25652429 - 33368333) on chromosome 6.
eFigure 2. Conditional quantile-quantile (Q-Q) plots of selected phenotypes as a function of Alzheimer’s Disease.

This supplementary material has been provided by the authors to give readers additional information about their work.
eAppendix 1. List of Investigators

Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of data and/or provided data but did not participate in analysis or writing of this report.

ADNI Investigators:


The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The following is a list of the consortia members:

IGAP consortium:

eAppendix 2. Additional Information

IGAP Cohort
International Genomics of Alzheimer's Project (IGAP) is a large two-stage study based upon genome-wide association studies (GWAS) on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 7,055,881 single nucleotide polymorphisms (SNPs) to meta-analyse four previously-published GWAS datasets consisting of 17,008 Alzheimer's disease cases and 37,154 controls (The European Alzheimer's disease Initiative – EADI the Alzheimer Disease Genetics Consortium – ADGC The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium – CHARGE The Genetic and Environmental Risk in AD consortium – GERAD). In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer's disease cases and 11,312 controls. Finally, a meta-analysis was performed combining results from stages 1 & 2. In this study, we focused on the stage 1 IGAP SNPs.

ADNI Cohort
The association between genotype and longitudinal clinical decline (ADAS-Cog) was assessed using data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date in formation, see www.adni-info.org.

Each participant was formally evaluated using eligibility criteria that are described in detail elsewhere (http://www.adni-info.org/index.php?option=com_content&task=view&id=9&Itemid=43). The institutional review boards of all participating institutions approved the procedures for this study. Written informed consent was obtained from all participants or surrogates. Experienced clinicians conducted independent semi-structured interviews with the participant and a knowledgeable collateral source that included a health history, neurological examination, and a comprehensive neuropsychological battery. We selected participants from the ADNI database if they were clinically diagnosed at baseline as cognitively normal (n = 196), amnestic mild cognitive impairment (MCI) (n = 355) or probable AD (n = 172).

Supplemental funding sources
IGAP: The i–Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG03193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728.
Conditional Q-Q plots

Q-Q plots compare a nominal probability distribution against an empirical distribution. In the presence of all null relationships, nominal p-values form a straight line on a Q-Q plot when plotted against the empirical distribution. For AD, CD, UC, RA, T1D, CeD, and PSOR SNPs and for each categorical subset (strata), \(-\log_{10} \text{nominal p-values}\) were plotted against \(-\log_{10} \text{empirical p-values}\) (conditional Q-Q plots, see Figure 1). Deflections of the observed distribution from the projected null line reflect increased tail probabilities in the distribution of test statistics (z-scores) and consequently an over-abundance of low p-values compared to that expected by chance (enrichment).

Under large-scale testing paradigms, such as GWAS, quantitative estimates of likely true associations can be estimated from the distributions of summary statistics (1, 2). One common method for visualizing the enrichment of statistical association relative to that expected under the global null hypothesis is through Q-Q plots of nominal p-values obtained from GWAS summary statistics. The usual Q-Q curve has as the y-ordinate the nominal p-value, denoted by “p”, and as the x-ordinate the corresponding value of the empirical cdf, denoted by “q”. Under the global null hypothesis the theoretical distribution is uniform on the interval \([0,1]\). As is common in GWAS, we instead plot \(-\log_{10} \text{p}\) against \(-\log_{10} \text{q}\) to emphasize tail probabilities of the theoretical and empirical distributions. Therefore, genetic enrichment results in a leftward shift in the Q-Q curve, corresponding to a larger fraction of SNPs with nominal \(-\log_{10} \text{p}\)-value greater than or equal to a given threshold. Conditional Q-Q plots are constructed by creating subsets of SNPs based on levels of an auxiliary measure for each SNP, and computing Q-Q plots separately for each level. If SNP enrichment is captured by variation in the auxiliary measure, this is expressed as successive leftward deflections in a conditional Q-Q plot as levels of the auxiliary measure increase.

We constructed conditional Q-Q plots of empirical quantiles of nominal \(-\log_{10}(\text{p})\) values for SNP association with AD for all SNPs, and for subsets (strata) of SNPs determined by the nominal p-values of their association with CD, UC, RA, T1D, CeD, and PSOR. Specifically, we computed the empirical cumulative distribution of nominal p-values for a given phenotype for all SNPs and for SNPs with significance levels below the indicated cut-offs for the other phenotypes (\(-\log_{10}(\text{p}) \geq 0, -\log_{10}(\text{p}) \geq 1, -\log_{10}(\text{p}) \geq 2\) corresponding to \(p < 1, p < 0.1, p < 0.01\) respectively). The nominal p-values (\(-\log_{10}(\text{p})\)) are plotted on the y-axis, and the empirical quantiles (\(-\log_{10}(\text{q})\), where \(q = 1 - \text{cdf}(\text{p})\)) are plotted on the x-axis (Figure 1). To assess for polygenic effects below the standard GWAS significance threshold, we focused the conditional Q-Q plots on SNPs with nominal \(-\log_{10}(\text{p}) < 7.3\) (corresponding to \(p > 5 \times 10^{-8}\)).

Genomic Control

The empirical null distribution in GWAS is affected by global variance inflation due to population stratification and cryptic relatedness (3) and deflation due to over-correction of test statistics for polygenic traits by standard genomic control methods (4). We applied a control method leveraging only intergenic SNPs, which are likely depleted for true associations (5). First, we annotated the SNPs to genic (5'UTR, exon, intron, 3'UTR) and intergenic regions using information from the 1KGP. We used intergenic SNPs because their relative depletion of associations suggests that they provide a robust estimate of true null effects and thus seem a better category for genomic control than all SNPs. We converted all p-values to z-scores and for all phenotypes we estimated the genomic inflation factor \(\lambda_{GC}\) for intergenic SNPs. We computed the inflation factor, \(\lambda_{GC}\) as the median z-score squared divided by the expected median of a chi-square distribution with one degree of freedom and divided all test statistics by \(\lambda_{GC}\).

Conditional True Discovery Rate (TDR)

Enrichment seen in the fold enrichment plots can be directly interpreted in terms of TDR (equivalent to one minus the False Discovery Rate (FDR)) (6). We applied the conditional FDR method (7), previously used for enrichment of GWAS based on linkage information (8). Specifically, for a given p-value cutoff, the FDR is defined as
\[
\text{FDR}(p) = \frac{\pi_0 F_0(p)}{F(p)},
\]
where \(\pi_0\) is the proportion of null SNPs, \(F_0\) is the null cdf, and \(F\) is the cdf of all SNPs, both null and non-null. Under the null hypothesis, \(F_0\) is the cdf of the uniform distribution on the unit interval \([0,1]\), so that Eq. \([1]\) reduces to
\[
\text{FDR}(p) = \frac{\pi_0}{F(p)},
\]
The cdf \(F\) can be estimated by the empirical cdf \(q = N_p / N\), where \(N_p\) is the number of SNPs with p-values less than or equal to \(p\), and \(N\) is the total number of SNPs. Replacing \(F\) by \(q\) in Eq. \([2]\), we get
\[
\text{Estimated FDR}(p) = \frac{\pi_0}{q},
\]
which is biased upwards as an estimate of the FDR (9). Replacing \(\pi_0\) in Equation \([3]\) with unity gives an estimated FDR that is further biased upward;
\[ q^* = \frac{p}{q} \quad [4] \]

If \( \pi_0 \) is close to one, as is likely true for most GWAS, the increase in bias from Eq. [3] is minimal. The quantity \( 1 - \frac{p}{q} \) is therefore biased downward, and hence is a conservative estimate of the TDR.

Referring to the formulation of the Q-Q plots, we see that \( q^* \) is equivalent to the nominal p-value divided by the empirical quantile, as defined earlier. Given the \( -\log_{10} \) of the Q-Q plots we can easily obtain

\[ -\log_{10}(q^*) = \log_{10}(q) - \log_{10}(p) \quad [5] \]

demonstrating that the (conservatively) estimated FDR is directly related to the horizontal shift of the curves in the conditional Q-Q plots from the expected line \( x = y \), with a larger shift corresponding to a smaller FDR, as illustrated in Figure 1. As before, the estimated TDR can be obtained as \( 1 - \text{FDR} \).

**Conjunction statistics – test of association with both phenotypes**

We defined the conjunction statistics (denoted as \( \text{FDR}_{\text{Trait1} \& \text{Trait2}} \)) as the maximum of the conditional FDR in both directions, i.e.

\[ \text{FDR}_{\text{Trait1} \& \text{Trait2}} = \max(\text{FDR}_{\text{Trait1} | \text{Trait2}}, \text{FDR}_{\text{Trait2} | \text{Trait1}}) \]

based on the combination of p-value for the SNP in AD and the associated immune-mediated disease (e.g. CD), by interpolation into a bidirectional 2-D look-up table (1, 2). The conjunction statistic allows for identification of SNPs that are associated with both phenotypes, which minimizes the effect of a single phenotype driving the common association signal. Table 1 lists all SNPs with conjunction FDR < 0.05 (-\( \log_{10}(\text{FDR}) > 1.3 \)) with AD and any of the immune-mediated diseases considered after removing all SNPs with \( r^2 > 0.2 \) based on 1KGP linkage disequilibrium (LD) (pruning).

**Conjunction FDR Manhattan plots**

To illustrate the localization of the genetic markers associated with AD given CD, UC, RA, T1D, CeD, and PSOR we used a ‘Conjunction FDR Manhattan plot’, plotting all SNPs within an LD block in relation to their chromosomal location. As illustrated in Figure 2 within the main manuscript, the large points represent the SNPs with FDR < 0.05, whereas the small points represent the non-significant SNPs. All SNPs before ‘pruning’ (removing all SNPs with \( r^2 > 0.2 \) based on 1KGP LD structure) are shown. The strongest signal in each LD block is illustrated with a black line around the circles. This was identified by ranking all SNPs in increasing order, based on the conditional FDR value for AD, and then removing SNPs in LD \( r^2 > 0.2 \) with any higher ranked SNP. Thus, the selected locus was the most significantly associated with AD in each LD block (Figure 2).

**Influence of the MHC region on genetic enrichment**

The major histocompatibility complex (MHC) is recognized as an important factor in the pathology of immune-mediated diseases. To examine the effect of the MHC on enrichment in AD SNPs, we evaluated the conditional Q-Q plots after removing the SNPs in the MHC region. We removed all SNPs located in the MHC (location 25652429 - 33368333) on chromosome 6 and all SNPs in LD with these MHC SNPs (\( r^2 > 0.2 \)) and re-examined the AD Q-Q plots conditional on the immune-mediated traits.
eReferences

eFigure 1. Conditional quantile-quantile (Q-Q) plots after removing all SNPs located in the MHC (location 25652429 - 33368333) on chromosome 6.

We plot empirical $-\log_{10} p$-values versus nominal $-\log_{10} p$-values (corrected for inflation) in Alzheimer’s disease (AD) below the standard GWAS threshold of $p < 5 \times 10^{-8}$ as a function of significance of association with Crohn’s disease (CD) (panel A), ulcerative colitis (UC) (panel B), Type 1 Diabetes (T1D) (panel C), rheumatoid arthritis (RA) (panel D), Celiac disease (CeD) (panel E), and psoriasis (PSOR) (panel F) at the level of $-\log_{10}(p) \geq 0$, $-\log_{10}(p) \geq 1$, $-\log_{10}(p) \geq 2$ corresponding to $p \leq 1$, $p \leq 0.1$, $p \leq 0.01$, respectively. Blue line indicates all SNPs.
eFigure 2. Conditional quantile-quantile (Q-Q) plots of selected phenotypes as a function of Alzheimer’s Disease.

Conditional quantile-quantile (Q-Q) plots of empirical $-\log_{10}$ p-values versus nominal $-\log_{10}$ p-values (corrected for inflation) in Crohn’s disease (CD) (panel A), ulcerative colitis (UC) (panel B), Type 1 Diabetes (T1D) (panel C), rheumatoid arthritis (RA) (panel D), Celiac disease (CeD) (panel E), and psoriasis (PSOR) (panel F) as a function of Alzheimer’s disease (AD) below the standard GWAS threshold of $p < 5 \times 10^{-8}$ as a function of significance of association with at the level of $-\log_{10}(p) \geq 0$, $-\log_{10}(p) \geq 1$, $-\log_{10}(p) \geq 2$ corresponding to $p \leq 1$, $p \leq 0.1$, $p \leq 0.01$, respectively. Blue line indicates all SNPs.