Implication of the APP Gene in Intellectual Abilities

Craig Myrum, Oleksii Nikolaienko, Clive R. Bramham, Jan Haavik and Tetyana Zayats

K.G. Jebsen Centre for Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway

Department of Psychiatry, Haukeland University Hospital, Bergen, Norway

Accepted 16 May 2017

Abstract

Background: Cognitive functions are highly heritable and polygenic, though the source of this genetic influence is unclear. On the neurobiological level, these functions rely on effective neuroplasticity, in which the activity-regulated cytoskeleton associated protein (ARC) plays an essential role.

Objectives: To examine whether the ARC gene complex may contribute to the genetic components of intellectual function given the crucial role of ARC in brain plasticity and memory formation.

Methods: The ARC complex was tested for association with intelligence (IQ) in children from the Avon Longitudinal Study of Parents and Children (ALSPAC, N = 5,165). As Alzheimer’s disease (AD) shares genetics with cognitive functioning, the association was followed up in an AD sample (17,008 cases, 37,154 controls).

Results: The ARC complex revealed association with verbal and total IQ (empirical p = 0.027 and 0.041, respectively) in the ALSPAC. The strongest single variant signal (rs2830077; empirical p = 0.018), within the APP gene, was confirmed in the AD sample (p = 2.76E-03). Functional analyses of this variant showed its preferential binding to the transcription factor CP2.

Discussion: This study implicates APP in childhood IQ. While follow-up studies are needed, this observation could help elucidate the etiology of disorders associated with cognitive dysfunction, such as AD.

Keywords: Alzheimer’s disease, APP, ARC, Avon Longitudinal Study of Parents and Children, dementia, intelligence, synaptic plasticity

INTRODUCTION

Cognitive function varies between individuals, and is a key predictor of important life outcomes such as mental and physical health as well as longevity [1]. General intellectual function, commonly measured by an aggregated score (intelligence quotient, IQ) across a wide range of cognitive tasks, can be considered a main trait behind this variation [2]. It is well established that intelligence is highly heritable and polygenic [3]. Nonetheless, the extent and nature of this genetic influence is still unknown [4] and warrants further investigation.

Intelligence captures a broad scope of cognitive abilities and can be enumerated by measures of verbal IQ (VIQ) and performance IQ (PIQ), often differentiated into crystallized and fluid types. Crystallized and fluid IQ represent different aspects of intelligence, both behaviorally and biologically [5]. Crystallized intelligence includes knowledge accumulated throughout life, determined by education and experience, while fluid intelligence consists of problem solving and reasoning abilities that has little reliance on stored knowledge [6]. Fluid intelligence is strongly correlated with working memory and functional activity observed by magnetic resonance imaging (MRI) during cognitive tasks [7, 8]. Although these two IQ measures can be combined into a full scale IQ (FSIQ), research on the neural basis of intelligence suggests that it could be best represented by at least two dimensions rather than one [5].
The heritability of intelligence has been estimated to increase with age, from ∼30% in childhood up to 80% in adulthood [9]. However, despite such high heritability, intelligence-related phenotypes have not been robustly associated with any single gene or variant [3, 10]. Only recently did a large-scale genome-wide association study (GWAS) on these traits identify 13 variants associated with general cognitive function [11]. This confirms that genetic architecture of intelligence involves many genes of small effect, suggesting that the impact of a single nucleotide polymorphism (SNP) may not be detected without large sample sizes.

The lack of robust association is common in the search for genes behind polygenic traits, such as intelligence [3, 11, 12]. Little progress has been made in finding loci that explain even a small fraction of the phenotypic variance in complex traits [13], posing one of the biggest challenges in present-day behavioral genetics. One promising approach to address this limitation is the examination of functionally related genes as a complex, thereby combining their effects [14].

At the neurobiological level, intellectual and cognitive functions rely on effective neuroplasticity, the brain’s ability to alter neural networks and synapses. Activity-regulated cytoskeleton associated protein (Arc) and its regulators and interactors have been extensively studied due to their essential role in multiple forms of synaptic plasticity [15–17]. Arc appears to play hub-like role in its regulation [18], at least partially by modifying actin cytoskeleton dynamics, receptor endocytosis, and glutamate receptor transcription [15–17]. Thus, ARC (here we refer to the gene as “ARC” and its RNA/protein as “Arc”) and its partners provide a promising set of candidate genes for uncovering genetic components of intellectual function. Here, we collectively refer to these functionally linked genes as the “ARC complex” (Table 1).

The genes in the ARC complex have been implicated in a number of conditions characterized by cognitive disabilities, including schizophrenia [19], Angelman syndrome [20], normal cognitive aging [21], and Alzheimer’s disease (AD) [22]. Notably, the ARC complex contains multiple AD-related genes, including amyloid precursor protein (APP), whose proteolysis generates an AD hallmark, amyloid-β (Aβ).

In this study, we hypothesized that the ARC complex may contribute to individual differences in intellectual function. We curated the complex and tested whether its SNPs are associated with IQ scores in children of the Avon Longitudinal Study of Parents and Children (ALSPAC). Since polygenic risk for AD is highly correlated with cognitive function [23], we next determined whether SNPs revealing significant association with intelligence in children were also associated with AD in adults in the International Genomics of Alzheimer’s Project (IGAP). Lastly, we examined whether the most strongly associated SNPs may have functional relevance in gene expression.

METHODS

Curation and Assessment of ARC Functional Complex

The ARC gene set was constructed based on Arc interaction partners and regulators, as evidenced in the literature (Table 1). We followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) Statement [24]. PubMed entries linked to human, mouse, and rat ARC genes (NCBI Gene IDs: 23237, 11838, and 54323, respectively) were obtained on December 14, 2014 using an NCBI cross-reference query. The search yielded 185 papers, all of which were screened to identify those papers showing experimental evidence for Arc interaction partners and regulators (n = 13). The remaining papers’ references were then examined further, which led to the identification of six additional relevant papers. From these 19 sources, a total of 37 experimentally-verified interaction partners and regulators were identified (Supplementary Figure 1). The gene complex was evaluated in STRING database, version 10 [25], to confirm its network connectivity and assess its enrichment for the biological and cellular processes based on gene ontology (GO). False discovery rate (FDR) was applied to correct for multiple testing.

Participants and genotyping

Avon Longitudinal Study of Parents and Children (ALSPAC) sample

The ALSPAC sample was derived from a well-characterized population-based study carried out in southwest England [26]. Pregnant women residing in the Bristol area of the United Kingdom, who had an estimated date of delivery between April 1, 1991 and December 31, 1992, were recruited to take part. Of the original 14,541 pregnancies, 13,988 children were alive at one year. An additional 713
Table 1

Summary of genes included in the curated ARC gene complex. The ARC gene complex was divided into two subgroups: (1) the “ARC expression” subgroup included proteins implicated in the regulation of ARC transcription, mRNA processing, transport, Arc protein translation and degradation and (2) the “ARC function” subgroup included proteins that bind Arc protein or are closely associated with Arc function.

<table>
<thead>
<tr>
<th>Entrez Gene ID</th>
<th>Gene Symbol</th>
<th>Level of interaction</th>
<th>Gene-Set subgroup</th>
<th>References (DOI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>23237</td>
<td>ARC</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1385</td>
<td>CREB1</td>
<td>DNA</td>
<td>ARC expression</td>
<td>10.1073/pnas.0806518106</td>
</tr>
<tr>
<td>4205</td>
<td>MEF2A</td>
<td>DNA</td>
<td>ARC expression</td>
<td>10.1073/pnas.0806518106</td>
</tr>
<tr>
<td>6722</td>
<td>SRF</td>
<td>DNA</td>
<td>ARC expression</td>
<td>10.1073/pnas.0806518106</td>
</tr>
<tr>
<td>3172</td>
<td>HNF4A</td>
<td>DNA</td>
<td>ARC expression</td>
<td>10.1073/pnas.0806518106</td>
</tr>
<tr>
<td>3182</td>
<td>HNRPAB</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1091/mfe.E10-11-0904</td>
</tr>
<tr>
<td>3181</td>
<td>HNRPAB2B1</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1091/mfe.E10-11-0904</td>
</tr>
<tr>
<td>5976</td>
<td>UPF1</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1016/j.cell.2007.05.028</td>
</tr>
<tr>
<td>9775</td>
<td>EIF4A3</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1016/j.cell.2007.05.028</td>
</tr>
<tr>
<td>407040</td>
<td>MIR34A</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1371/journal.pone.0041688</td>
</tr>
<tr>
<td>406968</td>
<td>MIR193A</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1371/journal.pone.0041688</td>
</tr>
<tr>
<td>442900</td>
<td>MIR326</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1371/journal.pone.0041688</td>
</tr>
<tr>
<td>1938</td>
<td>EIF2</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1371/journal.pone.0041688</td>
</tr>
<tr>
<td>2332</td>
<td>FMR1</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1016/j.cell.2008.07.031</td>
</tr>
<tr>
<td>8569</td>
<td>MKKN1</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1074/jbc.M109.056077</td>
</tr>
<tr>
<td>2872</td>
<td>MKKN2</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1074/jbc.M109.056077</td>
</tr>
<tr>
<td>23191</td>
<td>CYFIP1</td>
<td>mRNA</td>
<td>ARC expression</td>
<td>10.1016/j.cell.2008.07.031</td>
</tr>
<tr>
<td>6455</td>
<td>SH3GL1</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.neuron.2006.08.033</td>
</tr>
<tr>
<td>6457</td>
<td>SH3GL3</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.neuron.2006.08.033</td>
</tr>
<tr>
<td>6456</td>
<td>SH3GL2</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.neuron.2006.08.033</td>
</tr>
<tr>
<td>1785</td>
<td>DNM2</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.neuron.2006.08.033</td>
</tr>
<tr>
<td>816</td>
<td>CAMK2B</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.neuron.2006.08.033</td>
</tr>
<tr>
<td>1072</td>
<td>CFL1</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1523/NEUROSCI.2883-07.2007</td>
</tr>
<tr>
<td>4133</td>
<td>MAP2</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1002/jnr.20056</td>
</tr>
<tr>
<td>57731</td>
<td>SPTBN4</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.brainres.2007.03.079</td>
</tr>
<tr>
<td>1742</td>
<td>DLG4</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1371/journal.pbio.1001478</td>
</tr>
<tr>
<td>4915</td>
<td>NTRK2</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1371/journal.pbio.1001478</td>
</tr>
<tr>
<td>4851</td>
<td>NOTCH1</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.neuron.2011.01.004</td>
</tr>
<tr>
<td>351</td>
<td>APP</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.cell.2011.09.036</td>
</tr>
<tr>
<td>23621</td>
<td>BACE1</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.cell.2011.09.036</td>
</tr>
<tr>
<td>5663</td>
<td>PSEN1</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.cell.2011.09.036</td>
</tr>
<tr>
<td>7329</td>
<td>UBE2I</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1074/jbc.M112.356337</td>
</tr>
<tr>
<td>7337</td>
<td>UBE3A</td>
<td>protein</td>
<td>Arc expression</td>
<td>10.1016/j.cell.2010.01.026</td>
</tr>
<tr>
<td>54476</td>
<td>RNF216</td>
<td>protein</td>
<td>Arc expression</td>
<td>10.1016/j.neuron.2014.05.016</td>
</tr>
<tr>
<td>7341</td>
<td>SUMO1</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1074/jbc.M112.356337</td>
</tr>
<tr>
<td>5371</td>
<td>PML</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1016/j.brainres.2007.03.079</td>
</tr>
<tr>
<td>10524</td>
<td>KAT5</td>
<td>protein</td>
<td>Arc function</td>
<td>10.1523/NEURO.0019-14.2014</td>
</tr>
</tbody>
</table>

Children were enrolled after age seven, resulting in a total sample of 14,701 children. These mother-child pairs have been followed for over 20 years generating an immense amount of data through biological samples, measurements, and questionnaires. The study website contains details of all the data that is available through a fully searchable data dictionary: http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Phenotype-matched genotype data were available for up to 6,832 children, depending on the variables.
and prepared for genotyping using standard protocols. A total of 9,912 participants were genotyped using the Illumina HumanHap550 platform. SNPs with a minor allele frequency of <1% and call rate of <95% were removed. Furthermore, only SNPs passing the exact test of Hardy–Weinberg equilibrium ($p > 5.00 \times 10^{-7}$) were considered for analyses. Detailed quality control measures can be found at http://www.bristol.ac.uk/media-library/sites/alspac/migrated/documents/gwas-data-generation.pdf. Known autosomal variants were imputed with MACH 1.0.16 Markov Chain Haplotyping software [28], using CEPH individuals from phase 2 of the HapMap project (HG18) as a reference (release 22). Only SNPs with imputation quality estimates above 0.3 were examined.

Additional quality control was performed in the subset of individuals with available IQ measures to ensure that no SNPs or individuals had poor genotyping rates (<95%) and that no SNPs were rare (minor allele frequency, MAF <1%) or out of Hardy-Weinberg Equilibrium ($p < 1.00E-05$).

**International Genomics of Alzheimer’s Project (IGAP) sample**

ARC complex SNPs were extracted from the summary statistics of the IGAP GWAS (http://www.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php) on individuals of European ancestry. IGAP used genotyped and imputed data on 7,055,881 SNPs to meta-analyze four previously-published GWAS datasets [namely, 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), and the Genetic and Environmental Risk in AD consortium (GERAD)], consisting of 17,008 AD cases and 37,154 controls [29].

**Statistical analyses**

To examine the contribution of ARC complex to cognitive function, we carried out the following analyses: (1) tests for association between ARC complex genes and IQ scores in ALSPAC (gene-set tests); (2) genes found to be significant in the gene-set test were further examined for single SNP-based association in ALSPAC; and (3) given reported shared genetics between cognitive function and AD, significant SNPs in ALSPAC were examined in the IGAP to check for association with AD. These analyses are described in more detail below.

**Gene-set tests of ARC complex**

Here, we tested the entire ARC complex, followed by “ARC expression” and “Arc function” subgroups individually (Table 1) in ALSPAC. The “ARC expression” subgroup included proteins implicated in the regulation of ARC transcription, mRNA processing, transport, Arc protein translation and degradation, while the Arc function subgroup included proteins that bind Arc or are closely associated with Arc function. ARC itself was not considered as a member of “ARC expression” or “Arc function” subgroups as by examining those subgroups we aimed at elucidating the role of ARC regulators and interactors only. SNPs within each gene in the complex (including those within 5,000 base pairs upstream and 1,500 base pairs downstream of each gene) were analyzed in MAGMA, version 1.05 [30], where first the linkage disequilibrium (LD) adjusted gene-based $p$-values were calculated and then converted to Z-values to regress on gene-set membership, as a predictor. Gene size and gene-sets’ gene density were added as covariates to account for possible confounding effects. Only genes on autosomal chromosomes were examined. Bonferroni correction was applied to correct for multiple testing, with the significance threshold set to 4.17E-03 (12 tests).

**Single SNP-based tests**

Genes revealing significant association in the gene-set tests were further examined on a single SNP level. Linear regressions were performed to test for association between SNPs and ALSPAC IQ scores in PLINK, version 1.9 [31]. Sex was set as a covariate, and only SNPs on autosomal chromosomes were examined. Correction for multiple testing was achieved by one million permutations, producing a family-wise empirical $p$-value. SNPs reaching family-wise $p$-value below 0.05 were considered significant.

Variants found to be significantly associated with IQ after correction for multiple testing were then looked up in the IGAP to check for association with AD.

**Functional analyses**

**In silico assessment**

SNPs surviving correction for multiple testing in association analyses were examined in silico. We used RegulomeDB [32] and HaploReg (version 4.1; CEU population code; http://www.broadinstitute.org/mammals/haploreg/) [33] to explore whether any of the SNPs may affect gene expression. Given
the central role of hippocampus in ARC complex processes and memory formation, we focused on expression in hippocampal tissue. To investigate whether any of the SNPs are located within transcription factor binding sites, we utilized MatInspector (Matrix Library 9.3; http://www.genomatix.de/solutions/genomatix-genome-analyzer.html) and MATCH (version 1.0; http://www.gene-regulation.com/pub/programs.html) engines. Variants with predicted functional significance were then assessed further, using electrophoretic mobility shift assays (EMSA) and luciferase reporter assays.

**Electrophoretic mobility shift assay (EMSA)**

Synthetic 3′ biotin-labeled double-stranded oligonucleotides corresponding to the rs2830077[A] and rs2830077[C] sequences (20 fmol) and recombinant pure TFCP2 (100 ng, Active Motif) were incubated for 20 min using the GelshiftTM Chemiluminescent EMSA kit (Active Motif) in a 1 × binding buffer supplemented with 2.5% glycerol, 10 ng/μl Poly d(I-C), 0.1% NP-40, 50 mM KCl, 0.5 mM MgCl2, and 0.12 mM EDTA. Reaction mixtures were separated by 6% PAAG, and products were detected by streptavidin-HRP conjugate. For competition assays, unlabeled oligonucleotides at 100-fold molar excess were added to the reaction mixture 5 min before adding the biotin-labeled probe. Sequences of the double-stranded probes for rs2830077[A] and rs2830077[C] were 5′-GACACGCTGACTTCCAGGCACAAGAGACG-3′ and 5′-GACACGCTGACTTCCAGGCACAAGAGACG-3′, respectively.

**Luciferase reporter assay**

The 300 base pair region of APP intron 1, bearing rs2830077 (NC_000021.9: 26130583-26130284, GRCh38, p2 reference assembly) was amplified using Phusion® High-Fidelity DNA Polymerase (NEB) and specific primers tagged with XhoI restriction sites (forward: 5′-ATCCTCGAGTAGTTTCTAAAAACA TGG-3′ and reverse: 5′-ATCCTCGAGTAGTTTCTAAAAACA TGG-3′) and the A allele of rs2830077 was created by site-directed mutagenesis using QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent). The fragment was then excised from the pGEM-T Easy vector (Promega), and the A allele of rs2830077 was created by site-directed mutagenesis using QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent). The fragment was then excised from the pGEM-T Easy vector using XhoI restriction enzyme (NEB) and ligated into the same site within the pGL3-Basic vector (Promega) to ensure that the vector backbone used for all the promoter constructs remained the same. The sequences and orientation of inserts of all plasmids were verified by direct sequencing.

HEK293FT cells (Thermo Fisher Scientific) were cultured in DMEM (Sigma-Aldrich) supplemented with 10% FBS (Sigma-Aldrich), sodium pyruvate (1 mM), glutamine (2 mM), penicillin (100 U/ml) and streptomycin (100 μg/ml). Cells were plated into 96-well plates (1 × 104 cells/well) and co-transfected after 24 h using Lipofectamine 2000 (Invitrogen) with 80 ng pGL3 vector constructs and 20 ng pRL-TK control plasmid (Promega) to allow normalization of transfection efficiency. Transfections were carried out in at least triplicates and repeated six times in independent experiments. Cells were assayed for firefly and Renilla luciferase activities 48 h after transfection using the Dual-Luciferase® Reporter Assay System (Promega). Measurements were performed on a VICTOR3 Multilabel Plate Reader (Perkin Elmer). The ratio of firefly to Renilla luciferase activity was calculated.

**Image quantification and statistical analyses**

Densitometric analysis was done using NIH ImageJ (http://imagej.nih.gov/ij/) and its differences were assessed by one-way analyses of variance (ANOVA), using SigmaPlot 13.0 (Systat Software).

**RESULTS**

Curation and Assessment of ARC Functional Complex

After assembling our ARC complex (Table 1), we evaluated its functional network in STRING database, which uses experimental and predicted protein-protein interaction information to assess connectivity (Fig. 1). Protein-protein interactions were significantly enriched (p = 4.8E-14) in the full ARC complex, indicating that the proteins in the curated gene set are biologically connected as a group. The most significant enrichment of the ARC complex as a whole was noted for cellular components (neuron projection, GO:0043005, FDR = 2.59E-07 and a whole was noted for cellular components (neuron projection, GO:0043005, FDR = 2.59E-07 and a whole was noted for cellular components (neuron projection, GO:0043005, FDR = 2.59E-07 and a whole was noted for cellular components (neuron projection, GO:0043005, FDR = 2.59E-07 and a whole was noted for cellular components (neuron projection, GO:0043005, FDR = 2.59E-07) in the full ARC complex, indicating that the proteins in the curated gene set are biologically connected as a group. The most significant enrichment of the ARC complex as a whole was noted for cellular components (neuron projection, GO:0043005, FDR = 2.59E-07). Supplementary Table 1 reflects the main findings from these assessments.

Similar to ARC complex, STRING examination of its subgroups (“ARC expression” and “Arc function,” Table 1) revealed that both of them show indication of biological protein-protein interactions, with the “Arc function” subgroup displaying stronger evidence (p = 1.52E-09) than the “ARC expression”
Fig. 1. ARC complex network. Results from the STRING analysis.
Table 2
ALSPAC and IGAP participant characteristics and ALSPAC IQ scores

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of Participants</th>
<th>Males (%)</th>
<th>Age (years)</th>
<th>WISC-III (mean ± standard deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Verbal</td>
</tr>
<tr>
<td>ALSPAC</td>
<td>5,165</td>
<td>49.8</td>
<td>8.5</td>
<td>108.12 (16.56)</td>
</tr>
<tr>
<td>IGAP Cases</td>
<td>17,008</td>
<td>35.1–40.6</td>
<td>68.5–82.7*</td>
<td>NA</td>
</tr>
<tr>
<td>IGAP Controls</td>
<td>37,154</td>
<td>39.3–48.2</td>
<td>51.0–76.3**</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA, non-applicable. *Age range of time of onset across datasets. **Age range at time of examination across datasets.

subgroup ($p = 3.19E-04$). Correspondingly, the latter also showed less implication in GO pathways (Supplementary Figure 2) compared to the “Arc function” subgroup, where the strongest enrichment was observed for neuron projection (GO: 0043005, FDR = 1.47E-07), followed by protein binding (GO:0005515, FDR = 1.52E-07). Supplementary Figure 2 depicts the results of ARC complex subgroups STRING examination.

Participants and genotyping

The features of ALSPAC and IGAP samples are summarized in Table 2. The characteristics of IGAP sample were derived from its publication [29], combining 18 datasets. The distribution of the three IQ scores in ALSPAC is presented in Supplementary Figure 3.

Statistical analyses

Gene-set tests of ARC complex

Tentative association was observed between the functional subgroup (Table 1) of ARC complex and VIQ as well as FSIQ in the ALSPAC ($p = 0.027$ and 0.041, respectively). No association signal survived correction for multiple testing. Supplementary Table 2 shows the association results of all gene sets, and individual genes comprising those, for all IQ measures.

Single SNP-based tests

Since tentative association was observed for “ARC function” subgroup only, we tested individual SNPs in those genes for association with IQ measures in ALSPAC. The strongest signals were observed between 13 SNPs in the APP gene and PIQ, with 10 variants surviving the correction for multiple testing (Table 3, Supplementary Figure 4). Another strong signal was noted between rs1491579 in the SH3-domain GRB2-like 3 (SH3GL3) gene (Table 3) and FIQ, though it did not survive multiple test correction. Supplementary Table 3 shows results of association for all the variants reaching point-wise $p$-values below 1.00E-03.

The 10 APP SNPs surviving the correction for multiple testing in association analyses with the PIQ score were then examined for association with AD in IGAP. All 10 SNPs also revealed association signals with AD. Seven of these survived the Bonferroni correction ($p < 0.005$; Table 3 and Supplementary Figure 4).

Functional analyses

In silico assessment

Among the top APP SNPs significantly associated with PIQ, several were predicted to be located within active chromatin compartments, with transcription enhancer properties in hippocampus as well as expression quantitative trait loci (eQTL) activity in whole blood (Table 3).

H3K27Ac modification and DNaseI hypersensitivity data at rs2830077 also suggested that this region may be an enhancer (Supplementary Figure 5). The potential functional importance of this SNP was supported in HaploReg, which implicated rs2830077 in APP expression [34]. Further, RegulomeDB, MatInspector, and MATCH all indicated the presence of a putative transcription factor CP2 (TFCP2; also known as LBP-1c/CP2/LSF) binding site at rs2830077. We also found that the matrix score of rs2830077[A] is lower than that of the reference allele C, suggesting that the TFCP2 may have higher binding affinity to the C allele than to the A.

Electrophoretic mobility shift assay (EMSA)

To validate in silico predictions, we performed EMSA using pure recombinant TFCP2 and synthetic double-stranded oligonucleotides of the two alleles of rs2830077. EMSA demonstrated a shifted band of a DNA-protein complex, with a stronger intensity corresponding to its C allele (46% greater intensity of retarded band; Supplementary Figure 6), confirming its higher binding affinity of TFCP2.
Table 3
Properties of the SNPs (and those in high LD with them) revealing significant association between ARC complex variants and IQ measures in the ALSPAC sample

<table>
<thead>
<tr>
<th>SNP (reference allele)</th>
<th>Chr.</th>
<th>Base pair position (GRCh38)</th>
<th>Gene</th>
<th>Beta (95% CI)</th>
<th>Point-wise empirical p-value</th>
<th>Family-wise empirical p-value</th>
<th>OR (95% CI)</th>
<th>p-value in AD</th>
<th>r²</th>
<th>Chromatin state in hippocampus</th>
<th>Active chromatin marks in hippocampus</th>
<th>eQTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2830077 (C)</td>
<td>21</td>
<td>26130406</td>
<td>APP</td>
<td>1.40 (0.75–2.04)</td>
<td>2.00E-05</td>
<td>0.018</td>
<td>0.95 (0.93–0.98)</td>
<td>2.76E-03</td>
<td>1.00</td>
<td>Transcription Enhancer</td>
<td>H3K4me1_Enh</td>
<td>APP</td>
</tr>
<tr>
<td>rs432766 (C)</td>
<td>21</td>
<td>26135742</td>
<td>APP</td>
<td>1.40 (0.75–2.04)</td>
<td>2.00E-05</td>
<td>0.018</td>
<td>0.97 (0.94–0.99)</td>
<td>0.047</td>
<td>0.97</td>
<td>Transcription Enhancer</td>
<td>H3K4me1_Enh</td>
<td>APP</td>
</tr>
<tr>
<td>rs375369 (A)</td>
<td>21</td>
<td>26142055</td>
<td>APP</td>
<td>1.40 (0.75–2.04)</td>
<td>2.00E-05</td>
<td>0.018</td>
<td>0.96 (0.93–0.99)</td>
<td>0.015</td>
<td>0.94</td>
<td>none</td>
<td>H3K4me1_Enh</td>
<td>none</td>
</tr>
<tr>
<td>rs467021 (C)</td>
<td>21</td>
<td>26144759</td>
<td>APP</td>
<td>1.40 (0.75–2.04)</td>
<td>2.00E-05</td>
<td>0.018</td>
<td>0.95 (0.93–0.98)</td>
<td>3.06E-03</td>
<td>0.94</td>
<td>none</td>
<td>H3K4me1_Enh</td>
<td>none</td>
</tr>
<tr>
<td>rs12483003 (C)</td>
<td>21</td>
<td>26150590</td>
<td>APP</td>
<td>1.39 (0.74–2.03)</td>
<td>2.50E-05</td>
<td>0.020</td>
<td>0.95 (0.93–0.99)</td>
<td>3.53E-03</td>
<td>0.86</td>
<td>Transcription Enhancer</td>
<td>H3K4me1_Enh</td>
<td>none</td>
</tr>
<tr>
<td>rs368196 (C)</td>
<td>21</td>
<td>26158949</td>
<td>APP</td>
<td>1.39 (0.74–2.03)</td>
<td>2.50E-05</td>
<td>0.020</td>
<td>0.95 (0.92–0.98)</td>
<td>2.15E-03</td>
<td>0.91</td>
<td>Transcription Enhancer</td>
<td>H3K4me1_Enh</td>
<td>none</td>
</tr>
<tr>
<td>rs465235 (G)</td>
<td>21</td>
<td>26160704</td>
<td>APP</td>
<td>1.39 (0.74–2.03)</td>
<td>2.50E-05</td>
<td>0.020</td>
<td>0.95 (0.92–0.98)</td>
<td>1.27E-03</td>
<td>0.89</td>
<td>none</td>
<td>H3K4me1_Enh</td>
<td>none</td>
</tr>
<tr>
<td>rs2830072 (T)</td>
<td>21</td>
<td>26128343</td>
<td>APP</td>
<td>1.37 (0.72–2.02)</td>
<td>3.20E-05</td>
<td>0.029</td>
<td>0.96 (0.93–0.99)</td>
<td>0.020</td>
<td>0.98</td>
<td>Transcription Enhancer</td>
<td>H3K4me1_Enh</td>
<td>APP</td>
</tr>
<tr>
<td>rs2830073 (T)</td>
<td>21</td>
<td>26128825</td>
<td>APP</td>
<td>1.37 (0.72–2.02)</td>
<td>3.20E-05</td>
<td>0.029</td>
<td>0.95 (0.93–0.98)</td>
<td>2.47E-03</td>
<td>0.98</td>
<td>Transcription Enhancer</td>
<td>H3K4me1_Enh</td>
<td>APP</td>
</tr>
<tr>
<td>rs2154481 (C)</td>
<td>21</td>
<td>26101558</td>
<td>APP</td>
<td>1.36 (0.71–2.01)</td>
<td>3.50E-05</td>
<td>0.032</td>
<td>0.95 (0.92–0.98)</td>
<td>6.18E-04</td>
<td>0.84</td>
<td>Transcription Enhancer</td>
<td>H3K4me1_Enh</td>
<td>APP</td>
</tr>
<tr>
<td>rs2830066 (C)</td>
<td>21</td>
<td>26121885</td>
<td>APP</td>
<td>1.27 (0.62–1.92)</td>
<td>1.10E-04</td>
<td>0.096</td>
<td>NA</td>
<td>NA</td>
<td>0.80</td>
<td>Transcription Enhancer</td>
<td>H3K4me1_Enh</td>
<td>APP</td>
</tr>
<tr>
<td>rs2830088 (C)</td>
<td>21</td>
<td>26142422</td>
<td>APP</td>
<td>1.27 (0.62–1.92)</td>
<td>1.12E-04</td>
<td>0.098</td>
<td>NA</td>
<td>NA</td>
<td>0.83</td>
<td>none</td>
<td>H3K4me1_Enh</td>
<td>none</td>
</tr>
<tr>
<td>rs2830089 (G)</td>
<td>21</td>
<td>26142921</td>
<td>APP</td>
<td>1.27 (0.62–1.92)</td>
<td>1.12E-04</td>
<td>0.098</td>
<td>NA</td>
<td>NA</td>
<td>0.82</td>
<td>none</td>
<td>H3K4me1_Enh</td>
<td>none</td>
</tr>
<tr>
<td>rs1491579 (T)</td>
<td>15</td>
<td>38509927</td>
<td>SH3GL3</td>
<td>–1.26 (–1.89 – –0.64)</td>
<td>9.70E-05</td>
<td>0.060</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Active Enhancer</td>
<td>H3K4me1_Enh</td>
<td>SH3GL3</td>
</tr>
</tbody>
</table>

CI, confidence interval; Chr, chromosome; OR, odds ratio; NA, non applicable; LD, linkage disequilibrium; r², measure of linkage disequilibrium.
Luciferase reporter assay

To validate the ability of rs2830077 to modulate transcription, we performed a luciferase assay with the first APP intron harboring this SNP. The luciferase activities for both orientations of the A allele were lower than corresponding activities of the C allele (28% and 23% lower luciferase activity for antisense and sense orientation, respectively; Fig. 2), confirming that rs2830077 may alter enhancer-like activity of this region.

DISCUSSION

In this study, we utilized a pathway approach to study genetics of cognition in children (ALSPAC) and followed up the observed association signals in a sample of adults with and without AD (IGAP). Further, we performed a functional examination of associated variants in cell culture. Our experimental approach was prompted by evidence that: (1) cognitive ability may be influenced by genetic variation in synaptic plasticity-related genes [35], (2) AD shares polygenic etiology with cognitive functioning [23], and (3) pathway analyses of functionally related genes can be advantageous as they combine the effect of multiple genes that may be biologically meaningful [14]. Our analyses led us to uncover a common APP variant association with PIQ in children, which was replicated in AD in adults. We further found that this SNP could influence APP expression by affecting TFCP2 binding affinity.

APP encodes the amyloid-β protein precursor (AβPP) that forms Aβ-containing neuritic plaques, the accumulation of which is one of the key histopathological hallmarks in AD. While the role of APP in AD has long been known, its involvement in childhood intelligence has not been previously reported. The association signal observed in this study is between the APP and PIQ (Table 3), believed to reflect fluid intelligence. Examining the data from a recent large-scale GWAS on cognitive functions (N = 112,151) [36], we noted that APP gene revealed signs of association with reaction time (efficiency of information processing) and memory (p = 0.048 and 0.095, respectively; Supplementary Figure 7), further suggesting possible involvement of APP in cognition.

General fluid cognitive functioning in childhood has been proposed to be linked to late-onset dementia [37, 38], with genetic influences possibly being the driving force behind stability of cognitive functioning [39]. In a recent study, carriers of the APOE ε4 variant (the best replicated known genetic factor for AD) revealed correlation with working memory and attention in children [40]. Thus, we hypothesize that the genetics of PIQ may be an important determinant of cognitive abilities throughout the lifespan and of age-related dementia.

Genetic overlap between IQ and dementia has previously been reported in other studies, including a large genome-wide analysis in over 100,000 individuals [23]. Another GWAS on general fluid cognitive ability in adults (N = 53,949) identified four genes (namely TOMM40, APOE, ABCG1, and MEF2C) previously associated with AD or neuropathological features of AD and related dementias [11]. Such overlap has also been noted for the APP gene, where a coding variant was shown to be protective against both normal age-related cognitive decline and AD
While the APOE has shown robust association with AD and normal cognitive aging, more in-depth functional studies are needed to more thoroughly understand the functional significance of all genetic contributions of cognitive ability.

In this study, the most prominent significant signal was noted in APP, between the intronic rs2830077 and PIQ in children, also showing association with AD in adults (Table 3). In silico analyses of this SNP prompted its functional evaluation after noting that rs2830077: (1) is located within a region of active chromatin; (2) may have transcriptional enhancer activity in hippocampal tissue; (3) possesses eQTL activity; and (4) is located at a putative TFCP2 binding site. We confirmed via EMSA that TFCP2 binds to this putative enhancer region, with rs2830077 allele-specific binding affinity (Supplementary Figure 5). Further, luciferase assays indicate that the C allele of this SNP confers enhanced expression of APP (Fig. 2). The TFCP2 transcription factor has been implicated in erythroid gene expression [42], repression of HIV transcription [43], and in different cancer types [44, 45]. No direct connections between TFCP2 and AβPP expression have been reported, though a number of studies show association between TFCP2 gene and AD [46, 47]. The activity of TFCP2 itself is regulated by AβPP-interacting protein Fe65 [48] and the intracellular domain of APP-like protein 2 [49]. Another APP SNP (rs467021), in linkage disequilibrium with rs2830077 (r² = 0.99952), has been reported to be associated with cognitive decline in AD [50].

The dosage of AβPP has been implicated in altered neuronal endocytosis associated with increased Aβ production and age-related brain atrophy and degeneration, observed in patients with AD and Down syndrome (DS) [51]. Furthermore, duplication of the APP locus is thought to lead to early-onset AD [52] and the trisomy of this locus is likely to contribute to dementia in DS [53]. Indeed, the triplication of the Hsa21 segment including APP in people without DS has been associated with AD [53]. As both AD and DS display dementia, the observed association between PIQ and a SNP potentially altering the expression of APP, and thus its dosage, may help to elucidate the link between variation in intelligence and development of dementia.

We should note that to better estimate the contribution of genetics to IQ through life, examination of longitudinal samples where IQ is measured at early and late ages is desirable. However, since such samples were not available, we took the alternative approach of looking up significant child IQ association signals (ALSPAC) in late adulthood (IGAP). While our findings are intriguing, a direct replication in an independent childhood sample is warranted.

In summary, this study implicates APP in general cognitive abilities. We also show that evidence-based pathway analyses can be useful in identifying genetic factors underlying cognitive function. Follow-up studies are needed to more precisely determine how variants in APP may exert their effects on cognitive function over a lifespan. Such studies may have valuable implications for our understanding of etiology and, eventually, treatment of disorders associated with cognitive dysfunction, such as AD.

ACKNOWLEDGMENTS

We are grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and will serve as guarantors for the contents of this paper. This research was specifically funded by Stiftelsen Kristian Gerhard Jebsen, the University of Bergen, the Western Norwegian Health Authorities (Helse Vest), the Bergen Medical Research Foundation, and the European Community’s Seventh Framework Programme (FP7/2007–2013; Grant no 602805). GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

We thank the International Genomics of Alzheimer’s Project (IGAP) for providing summary results data for these analyses. 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 participants, the patients, and their families. The iSelect chips were 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 AG033193 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.

Finally, we thank professor Astri J. Lundervold for her invaluable insight on intelligence and help in reviewing this manuscript.

Authors’ disclosures available online (http://j-alz.com/manuscript-disclosures/17-0049r1).

**SUPPLEMENTARY MATERIAL**

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-170049.

**REFERENCES**


