Disease Markers 24 (2008) 293–309
IOS Press

Candidate gene discovery procedure after follow-up confirmatory analyses of candidate regions of interests for Alzheimer's disease in the NIMH sibling dataset

Tesfaye M. Baye[a,b,*], Rodney T. Perry[b], Howard W. Wiener[b], Zuomin Chen[b], Lindy E. Harrell[c] and Rodney C.P. Go[b]
[a]*Section on Statistical Genetics, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294, USA*
[b]*Department of Epidemiology and International Health, University of Alabama at Birmingham, Birmingham, AL 35294, USA*
[c]*Department of Neurology, University of Alabama at Birmingham, Birmingham, AL 35294, USA*

Abstract. The objective of this research was to develop a procedure to identify candidate genes under linkage peaks confirmed in a follow-up of candidate regions of interests (CRIs) identified in our original genome scan in the NIMH Alzheimer's diseases (AD) Initiative families (Blacker et al. [1]). There were six CRIs identified that met the threshold of multipoint lod score (MLS) of $\geqslant 2.0$ from the original scan. The most significant peak (MLS $= 7.7$) was at 19q13, which was attributed to *APOE*. The remaining CRIs with 'suggestive' evidence for linkage were identified at 9q22, 6q27, 14q22, 11q25, and 3p26. We have followed up and narrowed the 9q22 CRI signal using simple tandem repeat (STR) markers (Perry et al. [2]). In this confirmatory project, we have followed up the 6q27, 14q22, 11q25, and 3p26 CRIs with a total of 24 additional flanking STRs, reducing the mean interval marker distance (MID) in each CRI, and substantially increase in the information content (IC). The linkage signals at 6q27, 14q22 and 11q25 remain 'suggestive', indicating that these CRIs are promising and worthy of detailed fine mapping and assessment of candidate genes associated with AD.
We have developed a bioinformatics approach for identifying candidate genes in these confirmed regions based on the Gene Ontology terms that are annotated and enriched among the systematic meta-analyzed genes, confirmed by at least three case-control samples, and cataloged in the "AlzGene database" as potential Alzheimer disease susceptibility genes (http://www.alzgene.org).

Keywords: Alzheimer, linkage, QTL, STR, SNP, Genomic scan, Candidate gene, bioinformatics, gene ontology, GO, Alzforum, Alzgene database

1. Introduction

The most common form of dementia among aging people is Alzheimer's disease (AD), which involves the cerebral cortex and hippocampus which control thought, memory and language. Pathologically, AD is characterized by neurofibrillary tangles found in the neurons and the deposition of β-amyloid (Aβ) within senile plaques and cerebral blood vessels, resulting in the loss of neurons and synapses [3]. Clinically, AD is slowly progressive [4], usually beginning after the age of 65 years, and the risk increases with age, affecting 0.6% of the world's population for ages 65–69

*Corresponding author: Tesfaye M. Baye, Section on Statistical Genetics, Department of Biostatistics, RPHB 327, University of Alabama at Birmingham, Birmingham, AL 35294-0022, USA. Tel.: +1 205 975 9273; Fax: +1 205 975 2540; E-mail: tmersha@uab.edu.

ISSN 0278-0240/08/$17.00

years, but up to 22.2% of the population over 90 years of age [5]. The number of sufferers worldwide is estimated as over 20 million [6] with more than 5 million affected in the United States.

AD is clinically subdivided into early ($<$ 65 years) and late ($\geqslant$ 65 years) onset forms. About 10% of AD cases are familial with an autosomal dominant inheritance and these cases are often the early onset form [7]. Mutations in three genes, amyloid precursor protein (*APP*) on chromosome 21, presenilin 1 (*PS1*) on chromosome 14, and presenilin 2 (*PS2*) on chromosome 1 are estimated to account for about 50% of early-onset AD [8–10]. However, the majority of cases (90–95%) are late-onset AD (LOAD) that can show familial clustering without a clear Mendelian mode of inheritance [11]. Apolipoprotein E (*APOE*) on chromosome 19q13 has been confirmed by multiple independent studies [12] as a risk factor for LOAD, and is associated with lowering the age of onset [1]. However, *APOE* explains only 20–29% of the risk [13], and it is neither essential nor sufficient to cause AD [14–16]. The etiology of LOAD is complex with the possible involvement of several genes and/or environmental factors [17].

Efforts to identify additional LOAD loci have largely taken two main approaches: genome-wide linkage scans [1,18–21] and association studies of polymorphisms in candidate genes (for review see [22]). These studies indicate the existence of additional AD susceptibility genes on several chromosomes. In our genome-wide linkage scan of the National Institute of Mental Health Alzheimer's Disease Genetics Initiative (NIMH-ADGI) cohort of affected siblings, we identified five CRIs (9q22, 6q27, 14q22, 11q25, and 3p26) with suggestive linkage and one CRI on 19q23 that met criteria for 'significant' evidence of linkage defined by multipoint LOD scores (MLS) $\geqslant$ 2.0 [1]. Other linkage studies, though considerable overlap between samples from NIMH-ADGI sets [14], have identified overlapping regions or regions that are within a modest distance of three of these CRI on 19q23, 9q22, and 11q25 [15–17]. The 14q23 region was identified by an independent scan that used only Caribbean Hispanic samples [23], and serves as a separate replicate.

We followed up the CRI signal on chromosome 9q22 by genotyping additional simple tandem repeats (STRs) and found the region remained significant with an increase in the peak MLS from 2.9 to 3.8 at 95 cM and narrowing of the CRI to 11 cM (92–103 cM), thus supporting the region as potentially harboring LOAD genes [2]. In the present study, with the aim of confirming our previous whole-genome scan findings, we have conducted a follow-up study with denser STR markers spanning the four remaining CRIs (3p26, 6q27, 11q25, and 14q22). Based on the Gene Ontology (GO) terms that are annotated and enriched among the systematic meta-analyzed genes, confirmed by at least three case-control samples, and cataloged in the "AlzGene database" as potential Alzheimer disease susceptibility genes, we developed bioinformatics tools that extract potential AD candidate genes in these CRIs from genomic databases.

2. Materials and methods

2.1. Study population: NIMH AD genetic initiative families

The study subjects were collected as part of the NIMH Genetics Initiative following a standardized protocol utilizing the NINCDS-ADRDA criteria for diagnosis of definite, probable, and possible AD [24, 25]. A total of 468 families were ascertained. The primary structures of these families were affected sib-pairs [1], and the ethnic make-up was primarily Caucasian (95%). In 437 of these families, the mean age of onset (MAO) of affected family members was above 50 (mean = 72.4, range = 50–97) (Those with MAO $\leqslant$ 50 were believed to be enriched for the *APP*, *PS1*, and *PS2* mutations, and were therefore dropped from the current analyses). In addition to the total set (TS, $n = 437$ families), the linkage and mapping results presented here are also from a late age at onset subset with a MAO $\geqslant$ 65 (LOAD families) identified in 320 families and a subset of families identified as early-mixed (EM) with MAO between ages 51 and 65 ($n = 117$), both with similar primary structure and ethnic make-up as the total set of families. Blood was collected and sent to the NIMH repository at Rutgers University where genomic DNA was extracted from lymphocyte cell lines.

2.2. STR genotyping

STRs flanking the 6q27, 14q22, 11q25, 3p26 peaks were chosen from the genomic database (www.gdb.org) and deCODE Genetics (www.nature.com/ng/journal/v31/n3/ extref/ng917-S13.xls). The genetic locations of the STRs were based on sex average distances in the Rutgers combined linkage-physical map, Build 36.1 (compgen.rutgers.edu/mapomat). A total of 24 STRs were genotyped in an attempt to decrease the inter-

Table 1
Markers used in the study, and their positions on the Rutgers combined linkage- physical map of each chromosome

Chr. 6: Old CRI = 140–192 cM			Chr. 14: Old CRI = 40–80 cM			Chr. 11: Old CRI = 116–161 cM			Chr. 3: Old CRI = 0–44 cM		
Position cM	Original CIDR locus	Follow-up locus	Position cM	Original CIDR	Follow-up	Position cM	Original CIDR	Follow-up	Position cM	Original CIDR	Follow-up
140.2	D6S1009		37.4	D14S306		115.8	D11S1391		1.9	D3S2387	
148.3		D6S1003	48	D14S587		119.8		D11S965	6.0		D3S3630
150	GATA184A08		50.5		D14S276	122.6		D11S1885	13.2		D3S1560
153		D6S1637	56.6	D14S592		126.9	D11S1998		19.9	D3S1304	
156.1		D6S1687	59.3		D14S63	131.4		D11S4089	24.1	D3S4545	
158.2		D6S494	63.7	D14S588		137.2	D11S4464		31.6	DD3S1259	
160		D6S2420	68.7		D14S77	145.1	D11S912		36.9		D3S1286
162.9	D6S2436		69.8		D14S71	152.8		D11S4131	43.8	D3S3038	
165.8		D6S442	70.3		D14S43	160.7	D11S968				
169.4		D6S437	72.4		D14S61						
178.5	D6S1277		72.7	D14S53							
181.4		D6S1719	79.4	D14S606							
183		D6S264	84.5		D14S48						
188.5		D6S503									
191.6	D6S1027										

marker genomic distance in these four peak regions in these LOAD families, which was 9 cM in the original scan. The STRs genotyped for the original and follow-up mapping are listed in Table 1.

All STR genotyping for follow-up mapping on chromosomes 3p, 6q and 11q were performed with the Beckman/Coulter CEQ 8000 capillary electrophoresis platform with Well RED dyes (Beckman/Coulter, Fullerton, CA) as described previously [2]. The ratio of each of the flanking STRs in the two pooled products that were run on the CEQ platform was determined by the dye selection criteria [2]. The ratio of the six flanking markers on 6q27 (D6S1687, D6S2420, D6S1719, D6S494, D6S1637, D6S442) was 1.3:3:1.6:3:3.5:1.3, respectively while the ratio of flanking STRs on 3p26 and 11q25 (D3S3630, D3S1560, D3S1286, D11S965, D11S4089, D11S4131, D11S1885) were pooled as 3:2.75:1.25:2.75:3:1.8:1.25, respectively. The seven flanking STRs on chromosome 14q and the four remaining markers on chromosome 6q27 were genotyped by acrylamide electrophoresis using autoradiography as described previously [26].

2.3. Linkage analysis

The statistical analyses have been detailed previously [2]. In brief, model-free linkage analysis was performed using the program Genehunter Plus with extensions to calculate the Kong and Cox statistic [27,28]. Maximum likelihood estimates of allele frequencies were calculated with the SAGE [29] program FREQ taking into account the family relationships. Replicates were performed on selected samples and any Mendelian errors were detected with the SAGE program MARKERINFO, as well as detected implicitly by the analytical programs used here.

2.4. Bioinformatics approach for candidate gene selection in CRIs

Once a linkage peak has been identified, hundreds of genes under the peak can be accessed using the UCSC genome browser (http://genome.ucsc.edu). But the lists are too large to conduct expensive molecular lab work. We developed a bioinformatics approach using a Python (http://www.python.org) script to assist in the efficient automatic extraction of candidate genes in regions of linkage (the codes are available upon request). Specific procedures for AD are: 1. The region to be analyzed needs to be defined. The input regions are usually defined as the 1 LOD drop of the linkage peak whose physical location is defined by the closest markers to the ends of the regions. 2. A list of disease specific keywords generated from 24 meta-analyzed and confirmed AD genes (http://www.alzgene.org; accessed on July 2, 2007) is provided to the program. These lists are function and process key terms deposited in the Gene Ontology (GO) databases (http://www.geneontology.org; accessed on July 2, 2007) for this disease. GO is developed to capture the activities at cellular and molecular level. For AD, these terms are, in general, related to neurological function, inflammation, oxidative damage, cholesterol/lipoprotein function and atherosclerosis/vascular pathways [30]. As of 2006, there are about 18,455 GO terms assigned to proteins to illustrate what they do, where they function, and what processes they are involved in. The AlzGene (http://www.alzgene.org) with in the Alzforum (http://www.alzforum.org) cat-

alogue, "AlzGene database", is a systematic, meta-analyses of potential Alzheimer disease susceptibility genes in at least three case-control samples [22]. This gene database is expected to provide a powerful tool for deciphering the genetics of Alzheimer disease, and serve as a potential model for tracking the most viable candidate genes.

Based on the GO vocabulary terms of the 24 most significant genes identified by Algene, we developed a bioinformatics approach to identify candidate genes in our QTL regions linked to AD. Briefly, we articulate that these GO terms for these 24 candidate liability genes (Table 2) can serve as a model in discovering new AD genes in each of our linkage peak regions. For each gene, we downloaded two categories of GO terms, function and process, from GO databases and interrogated the linkage regions of interest with theses GO terms for possible candidate AD genes. Genes enriched with the specified AD related GO terms are then downloaded directly from the human genome draft Build 36.1 database (http://genome.ucsc.edu) assembly.

3. Results

3.1. CRI located at 6q27

The chromosome 6q27 interval (140–192 cM) had a peak MLS score of 2.2. We have genotyped an additional ten markers and decreased the MID in this region from 10.9 cM to 3.8 cM. Additionally, the interval of support in the follow-up scan narrows to 148–188 cM (Table 1). Graph of the MLS scores from the total dataset showed the extra marker information split the peak: one signal located between 140 and 165 cM in the LOAD group with a peak MLS of 0.95 at 153 cM and a second CRI located between 170 and 192 cM in the EM group with a peak MLS of 1.48 at 183 cM (Fig. 1a).

3.2. CRI located at 14q22

The CRI we identified at 14q22 (37–80 cM) in the original scan had the highest MLS in the EM subset of families (MLS = 2.2). This evidence is probably due in large part to the presence of the PSEN1 gene located within this region at 68.7 cM. We have genotyped an additional 7 markers in this region, reducing the MID from 6.7 to 3.0 cM, and increasing the IC from 0.30–0.54 to 0.44–0.65. Additionally, the interval of support in the follow-up scan narrows to 50–85 cM (Table 1). The CRI is confirmed in the EM group between 35 and 75 cM with an increase in the peak MLS score from 2.2 to 2.5 at 48 cM (Fig. 1b). This places the peak ∼20 cM proximal to *PSEN1*, which suggests the location of a possible second susceptibility gene in the EM group.

3.3. CRI located at 11q27

Four additional microsatellites were genotyped in the CRI located at 11q25 (116–161 cM), resulting in a decrease in the MID of this region to 5.0 cM from an MID of 9.0 cM, and an increase of IC from 0.39–0.52 to 0.45–0.62. Additionally, the interval of support in the follow-up scan narrows to 120–153 cM (Table 1). The peak MLS remained unchanged from the original scan (2.0 at 158 cM) for the total set and there was narrowing of the signal to a region between 130 and 164 cM (Fig. 1c).

3.4. CRI located at 3p26

The CRI located at 3p26 had an MLS of 2.0 in the original scan in the EM and CM sets. We have genotyped three additional markers in this region, that result an increase in IC from 0.32–0.52 to 0.43–0.65 and a decrease in the MID from 8.8 to 5.3 cM. The peak MLS for the EM set and the combined set decreased to 1.45 and there was no narrowing of the region. Since we wish to maintain our stringent criteria of an MLS $\geqslant$ 2.0 for confirmation of a CRI and because there is no supporting evidence of the 3p26 region linked to AD in other genomic scans, this signal is most likely a false positive, therefore we have omitted this region from any further investigations.

3.5. Candidate gene selection

Results of possible candidate genes in these CRIs and their locations from the interrogation of the database with the GO terms are presented on Table 3. Approximately 159 position-based candidate genes and nearly 53,313 HapMap Caucasian SNPs (http://genome.ucsc.edu) were found at the 6q27 CRI (140–192 cM peak). Our bioinformatics procedure reduced the candidate gene lists to 56 based on function and 48 based on process, with 24 that were overlapping genes. At the interval (35–75 cM) of the 14q22 peak, about 198 genes and nearly 58,438 HapMap Caucasian SNPs were found. The interrogation of the database using GO terms reduced the candidate genes to 107 based on function and 69 based on process, with 46 that

Table 2
Systematic Meta-analyzed AD genes with Function and Process Gene Ontology Terms

No	Gene	Name	Position	GO terms	
				Function	Process
1	*APOE*	apolipoprotein E precursor	19q13.2	antioxidant activity apolipoprotein E receptor binding beta-amyloid binding heparin binding lipid transporter activity lipoprotein binding phospholipid binding tau protein binding	calcium ion homeostasis cholesterol catabolic process cholesterol homeostasis circulation cytoskeleton organization and biogenesis induction of apoptosis intracellular transport learning and/or memory lipid transport lipoprotein metabolic process protein tetramerization regulation of axon extension regulation of neuronal synaptic plasticity response to reactive oxygen species synaptic transmission, cholinergic vasodilation
2	*CHRNB2*	cholinergic receptor, nicotinic, beta polypeptide 2 (neuronal)	1q21.3	acetylcholine receptor activity extracellular ligand-gated ion channel activity ion channel activity nicotinic acetylcholine-activated cation-selective channel activity	ion transport memory sensory perception signal transduction
3	*CH25H*	cholesterol 25-hydroxylase	10q23	iron ion binding metal ion binding steroid hydroxylase activity	cholesterol metabolic process lipid metabolic process metabolic process sterol biosynthetic process
4	*PGBD1*	piggyBac transposable element derived 1	6p22.1	scavenger receptor activity transcription factor activity	regulation of transcription, DNA-dependent
5	*LMNA*	lamin A/C isoform 3	1q21.2-q21.3	protein binding structural molecule activity	
6	*SOAT1*	sterol O-acyltransferase (acyl-Coenzyme A: cholesterol acyltransferase) 1	1q25	acyltransferase activity acyltransferase activity sterol O-acyltransferase activity transferase activity	cholesterol metabolic process circulation lipid metabolic process steroid metabolic process
7	*MAPT*	microtubule-associated protein tau isoform 4	17q21.1	enzyme binding lipoprotein binding microtubule binding microtubule binding	generation of neurons generation of neurons microtubule cytoskeleton organization and biogenesis microtubule cytoskeleton organization and biogenesis

Table 2, continued

No	Gene	Name	Position	GO terms	
				Function	Process
				SH3 domain binding	negative regulation of microtubule depolymerization
				structural constituent of cytoskeleton	positive regulation of axon extension
				structural constituent of cytoskeleton	positive regulation of axon extension positive regulation of microtubule polymerization positive regulation of microtubule polymerization
8	*SORL1*	sortilin-related receptor, L(DLR class) A repeats-containing	11q23.2-q24.2	lipid transporter activity	cholesterol metabolic process
				transmembrane receptor activity	lipid metabolic process lipid transport receptor-mediated endocytosis steroid metabolic process
9	*PCK1*	cytosolic phosphoenolpyruvate carboxykinase 1	20q13.31	GTP binding	gluconeogenesis
				lyase activity	glycerol biosynthetic process from pyruvate
				nucleotide binding	lipid metabolic process
				phosphoenolpyruvate carboxykinase (GTP) activity	
10	*CST3*	cystatin C (amyloid angiopathy and cerebral hemorrhage)	20p11.21	cysteine protease inhibitor activity	
				protein homodimerization activity	
11	*ACE*	angiotensin I converting enzyme isoform 2	17q23.3	carboxypeptidase activity IEA	blood pressure regulation
				chloride ion binding	metabolic process
				hydrolase activity, acting on glycosyl bonds	proteolysis
				metal ion binding	proteolysis
				peptidyl-dipeptidase A activity	
				peptidyl-dipeptidase A activity	
				zinc ion binding	
12	*SORCS1*	sortilin-related VPS10 domain containing receptor 1	10q23-q25	neuropeptide receptor activity	neuropeptide signaling pathway
				protein binding	
				protein binding	
13	*TF*	transferrin	3q22.1	ferric iron binding	ion transport
				metal ion binding	iron ion homeostasis
					iron ion transport
14	*GALP*	galanin-like peptide precursor	19q13.42	neuropeptide hormone activity	biological process neuropeptide signaling pathway
15	*CTSD*	cathepsin D (lysosomal aspartyl protease)	11p15.5	cathepsin D activity	proteolysis
				pepsin A activity	
				peptidase activity	

Table 2, continued

No	Gene	Name	Position	GO terms	
				Function	Process
16	*TNK1*	tyrosine kinase, non-receptor, 1	17p13.1	ATP binding non-membrane spanning protein tyrosine kinase activity nucleotide binding protein binding signal transducer activity transferase activity	protein amino acid autophosphorylation
17	*IL1B*	interleukin 1, beta proprotein	2q14	interleukin-1 receptor binding protein binding signal transducer activity	apoptosis cell proliferation IEA cell-cell signaling fever immune response inflammatory response leukocyte migration negative regulation of cell proliferation neutrophil chemotaxis positive regulation of chemokine biosynthetic process positive regulation of interleukin-6 biosynthetic process regulation of progression through cell cycle signal transduction
18	*PON1*	paraoxonase 1	7q21.3	aryldialkylphosphatase activity arylesterase activity hydrolase activity lipid transporter activity	response to external stimulus response to toxin
19	*DAPK1*	death-associated protein kinase 1	9q34.1	ATP binding calmodulin binding kinase activity nucleotide binding protein serine/threonine kinase activity transferase activity	apoptosis induction of apoptosis by extracellular signals protein amino acid phosphorylation protein kinase cascade signal transduction
20	*PRNP*	prion protein interacting protein	1p32	exonuclease activity hydrolase activity	
21	*MYH13*	myosin, heavy polypeptide 13, skeletal muscle	17p13	actin binding ATP binding calmodulin binding microfilament motor activity nucleotide binding	striated muscle contraction
22	*HMGCS2*	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	1p13-p12	hydroxymethylglutaryl-CoA synthase activity transferase activity	acetyl-CoA metabolic process cholesterol biosynthetic process

Table 2, continued

No	Gene	Name	Position	GO terms	
				Function	Process
23	*BDNF*	brain-derived neurotrophic factor isoform b	11p13	growth factor activity growth factor activity protein binding	anti-apoptosis axon guidance axon target recognition dendrite development feeding behavior glutamate secretion inner ear development mechanoreceptor differentiation negative regulation of neuroblast proliferation nerve development neuron recognition positive regulation of neuron differentiation regulation of metabolic process regulation of neuron apoptosis regulation of retinal programmed cell death regulation of synaptic plasticity ureteric bud development
24	*PSEN1*	presenilin 1 isoform I-463	14q24.3	peptidase activity protein binding	amyloid precursor protein catabolic process anti-apoptosis apoptosis cell adhesion chromosome organization and biogenesis (sensu Eukaryota) chromosome segregation intracellular signaling cascade membrane protein ectodomain proteolysis Notch receptor processing positive regulation of enzyme activity protein processing regulation of phosphorylation

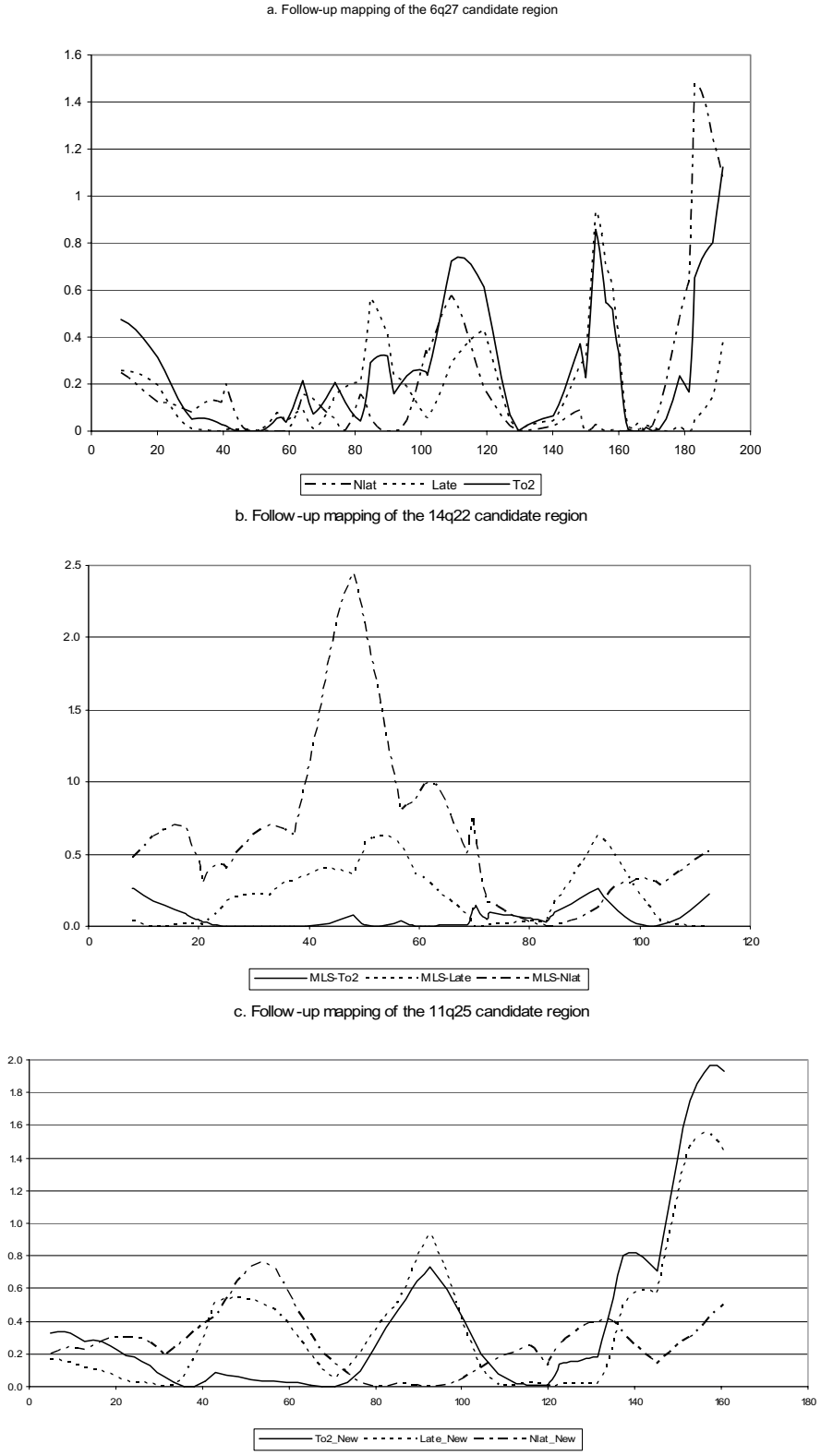

Fig. 1. MLS data of follow-up scan with an additional dense grid of markers for the CRIs on chromosome 6 (1a), 14 (1b), 11 (1c).

Table 3
Lists of total, functional, process genes in 6q, 14q and 11q CRI

chr 6 (140–192cM)			chr 14 (35–75cM)			chr 11(130–164cM)		
Total 159 genes	Function 56 genes	Process 48 genes	Total 198 genes	Function 107 genes	Process 69 genes	Total 83 genes	Function 12 genes	Process 26 genes
ACAT2	ACAT2	**AKAP12**	ABCD4	ABCD4	**ADAM20**	AB045830	ARHGEF12	**ESAM**
AGPAT4	AGPAT4	**ARID1B**	ACTN1	ACOT1	**ADAM21**	ACAD8	ESAM	**GRIK4**
AIG1	AKAP12	**BCLAF1**	ACTR10	ACOT2	**AKAP5**	ACRV1	GRIK4	**NRGN**
AKAP12	ARID1B	C6orf103	ACYP1	ACOT4	**ARF6**	ADAMTS15	HSPA8	OR10G4
ARID1B	BCLAF1	C6orf123	ADAM20	ACTN1	ATP5S	ADAMTS8	NRGN	OR10G7
BCLAF1	CITED2	C6orf54	ADAM21	ACTR10	**ATP6V1D**	APLP2	PKNOX2	OR10G8
BRP44L	CLDN20	CCR6	ADCK1	ACYP1	**BATF**	ARHGEF12	POU2F3	OR10G9
C6orf103	CNKSR3	**ESR1**	AF161445	ADAM20	**BMP4**	ASAM	SC5DL	OR4D5
C6orf115	EPM2A	**FBXO5**	AF336880	ADAM21	CDKL1	AY358331	SIAE	OR6M1
C6orf118	ESR1	GPR126	AF390028	AKAP5	**CDKN3**	B3GAT1	SORL1*	OR6X1
C6orf120	FBXO30	GPR31	AF390029	ARF6	**CGRRF1**	BARX2	TRIM29	OR8B12
C6orf122	FBXO5	GRM1	AF435011	ARG2	**CHURC1**	BC030580	ZNF202	OR8B2
C6orf123	HIVEP2	GTF2H5	AHSA1	ARID4A	**CHX10**	BRCC2		OR8B3
C6orf208	KATNA1	**HIVEP2**	AKAP5	ATP6V1D	CNIH	CDON		OR8B4
C6orf211	KIF25	IFNGR1	AL833179	BATF	**DLG7**	CHEK1		OR8B8
C6orf35	LATS1	IGF2R	ALDH6A1	BMP4	**DPF3**	CR625776		OR8D1
C6orf54	LPA	**LATS1**	ALKBH	BTBD5	ERH	CRTAM		OR8D2
C6orf55	MAP3K5	LOC653483	ARF6	C14orf4	**ESR2**	CSE-C		OR8D4
C6orf59	MAP3K7IP2	**LPA**	ARG2	CDKL1	**ESRRB**	DCPS		OR8G2
C6orf70	MAP7	LPAL2	ARID4A	CDKN3	**FOS**	DDX25		OR8G5
C6orf71	MLLT4	MAP3K4	ATP6V1D	CGRRF1	**FOXA1**	ESAM		**PKNOX2**
C6orf72	MTHFD1L	**MAP3K5**	AY451390	CHURC1	**GNG2**	FEZ1		ROBO3
C6orf96	PARK2	MAS1	BATF	CHX10	GPR135	FLI1		SCN3B
C6orf97	PCMT1	**MLLT4**	BC006221	CTAGE5	**HIF1A**	GRIK4		**SORL1***
CCDC28A	PDE10A	NMBR	BC011548	DAAM1	**KCNH5**	HSPA8		SPA17
CCR6	PDE7B	OLIG3	BC090945	DDHD1	LTBP2	ITM1		**ZNF202**
CITED2	PERP	OPRM1	BMP4	DLG7	MAP3K9	JAM3		
CLDN20	PEX3	**PDE10A**	BRMS1L	DLST	**MAP4K5**	KCNJ1		
CNKSR3	PHACTR2	**PDE7B**	BTBD5	DPF3	**MAX**	KCNJ5		
DACT2	PHF10	**PERP**	C14orf101	ENTPD5	**MNAT1**	KIRREL3		
DEADC1	PLAGL1	**PHF10**	C14orf106	ESR2	**NEK9**	LOH11CR2A		
DLL1	PLG	**PLAGL1**	C14orf111	ESRRB	NID2	LRRC35		
DYNLT1	PPIL4	**PLG**	C14orf140	EXDL2	**NKX2-8**	NFRKB		
EPM2A	RAB32	**RAB32**	C14orf162	FANCM	NPC2	NRGN		
ESR1	RAET1E	RPS6KA2	C14orf166	FNTB	**NRXN3**	OPCML		
FAM120B	RAET1G	**SHPRH**	C14orf29	FOS	**OTX2**	OR10G4		
FAM54A	RBM16	**SLC22A1**	C14orf31	FOXA1	**PAPLN**	OR10G7		
FBXO30	RGS17	SLC22A2	C14orf32	GCH1	PAX9	OR10G8		
FBXO5	RNASET2	SLC22A3	C14orf4	GMFB	**PGF**	OR10G9		
FGFR1OP	SHPRH	**SNX9**	CDKN3	GNG2	PLEK2	OR10S1		
FLJ27255	SLC22A1	TAGAP	CGRRF1	GNPNAT1	**PLEKHC1**	OR4D5		
FLJ37060	SLC22A2	**THBS2**	CHURC1	GSTZ1	PNN	OR6M1		
FLJ39824	SNX9	TIAM2	CHX10	HIF1A	PPP2R5E	OR6T1		
FLJ44955	SOD2	**TNFAIP3**	CLEC14A	HSPA2	**PRKCH**	OR6X1		
FNDC1	STX11	**TULP4**	CNIH	KCNH5	**PSEN1***	OR8A1		
FRMD1	SYNE1	UNC93A	COQ6	MAP3K9	PTGDR	OR8B12		
FUCA2	SYNJ2	VIP	CTAGE5	MAP4K5	PTGER2	OR8B2		
GPR126	SYTL3	**ZBTB2**	DAAM1	MAX	**RAB15**	OR8B3		
GPR178	TCP1		DACT1	MLH3	**RGS6**	OR8B4		
GPR31	THBS2		DDHD1	MNAT1	**RPS6KL1**	OR8B8		
GRM1	TNFAIP3		DHRS7	MPP5	**RTN1**	OR8D1		
GTF2H5	TULP4		DLG7	MTHFD1	SAV1	OR8D2		
HEBP2	UST		DLST	NEK9	**SIX1**	OR8D4		
HECA	VIL2		EIF2B2	NGB	**SIX4**	OR8G2		
HIVEP2	ZBTB2		EIF2S1	NIN	**SIX6**	OR8G5		
IFNGR1	ZDHHC14		ENTPD5	NKX2-8	SLC10A1	PANX3		
IGF2R			ERH	NRXN3	**SLC8A3**	PATE		

Table 3, continued

chr 6 (140–192cM)			chr 14 (35–75cM)			chr 11(130–164cM)		
Total 159 genes	Function 56 genes	Process 48 genes	Total 198 genes	Function 107 genes	Process 69 genes	Total 83 genes	Function 12 genes	Process 26 genes
IL20RA			ERO1L	NUMB	SNAPC1	PMP22CD		
IL22RA2			ESR2	OTX2	SOCS4	POU2F3		
KATNA1			FBXO33	PAPLN	**SPG3A**	PRDM10		
KIAA1244			FBXO34	PELI2	SSTR1	PUS3		
KIF25			FKBP3	PGF	**TGFB3**	RICS		
LATS1			FNTB	PLEKHC1	**TITF1**	ROBO3		
LOC202459			FOS	PNMA1	TXNDC	ROBO4		
LOC401280			FOXA1	POLE2	VTI1B	RPUSD4		
LOC401286			FUT8	PPM1A	WDHD1	SC5DL		
LOC441177			GARNL1	PRKCH	ZBTB1	SCN3B		
LOC441179			GCH1	PSEN1*	**ZBTB25**	SLC37A2		
LOC653483			GMFB	PSMA3	**ZNF410**	SNX19		
LPA			GNG2	PSMC6		SORL1		
LPAL2			GNPNAT1	PYGL		SPA17		
LRP11			GPHN	RAB15		SPATA19		
LTV1			GPR135	RAD51L1		SRPR		
MAP3K4			GPX2	RBM25		ST14		
MAP3K5			GSTZ1	RGS6		ST3GAL4		
MAP3K7IP2			HBLD1	RHOJ		STS-1		
MAP7			HIF1A	RPS6KL1		TBRG1		
MAS1			HSPA2	RTN1		TECTA		
MGC35308			JDP2	SAV1		THY28		
MLLT4			KCNH5	SEC23A		TIRAP		
MRPL18			KIAA0317	SFRS5		TMEM45B		
MTHFD1L			KIAA0586	SGPP1		VSIG2		
MTRF1L			KIAA1036	SIP1		ZNF202		
MYCT1			KIAA1393	SIPA1L1				
NMBR			KIAA1737	SIX1				
NOX3			KLHDC1	SIX4				
NUP43			KLHDC2	SIX6				
OLIG3			KTN1	SLC8A3				
OPRM1			LGALS3	SNW1				
PACRG			LRFN5	SPG3A				
PARK2			LTBP2	SPTB				
PBOV1			MAMDC1	SPTLC2				
PCMT1			MAP3K9	STYX				
PDE10A			MAP4K5	SYNE2				
PDE7B			MAX	SYNJ2BP				
PERP			MBIP	TGFB3				
PEX3			MED6	TIMM9				
PEX7			MGAT2	TITF1				
PHACTR2			MIA2	TRIM9				
PHF10			MIPOL1	TTLL5				
PIP3-E			MLH3	ZADH1				
PLAGL1			MNAT1	ZBTB1				
PLG			MPP5	ZBTB25				
PNLDC1			MTHFD1	ZFP36L1				
PPIL4			NEK9	ZFYVE1				
PPP1R14C			NGB	ZFYVE26				
QKI			NID2	ZNF410				
RAB32			NIN					
RAET1E			NKX2-8					
RAET1G			NPC2					
RAET1L			NRXN3					
RBM16			NUMB					
REPS1			OTX2					
RGS17			PAPLN					
RNASET2			PAX9					
RPS6KA2			PCNX					

Table 3, continued

chr 6 (140–192cM)			chr 14 (35–75cM)			chr 11(130–164cM)		
Total 159 genes	Function 56 genes	Process 48 genes	Total 198 genes	Function 107 genes	Process 69 genes	Total 83 genes	Function 12 genes	Process 26 genes
RSHL2			PELI2					
SASH1			PGF					
SERAC1			PIGH					
SF3B5			PLEK2					
SFT2D1			PLEKHC1					
SHPRH			PLEKHG3					
SLC22A1			PLEKHH1					
SLC22A2			PNMA1					
SLC22A3			PNN					
SLC35D3			POLE2					
SMOC2			POMT2					
SNX9			PPIL5					
SOD2			PPM1A					
STX11			PPP2R5E					
STXBP5			PRKCH					
SUMO4			PRPF39					
SYNE1			PSEN1					
SYNJ2			PSMA3					
SYTL3			PSMC6					
T			PTGDR					
TAGAP			PTGER2					
TCP1			PYGL					
TCP10			RAB15					
TCP10L2			RAD51L1					
TCTE3			RBM25					
TFB1M			RDH11					
THBS2			RDH12					
TIAM2			RGS6					
TNFAIP3			RHOJ					
TTLL2			RPL10L					
TULP4			RPL36AL					
TXLNB			RPS29					
ULBP1			RPS6KL1					
ULBP2			RTN1					
ULBP3			SAMD4					
UNC93A			SAV1					
UST			SDCCAG1					
VIL2			SEC10L1					
VIP			SEC23A					
WDR27			SFRS5					
WTAP			SGPP1					
ZBTB2			SIP1					
ZDHHC14			SIPA1L1					
			SIX1					
			SIX4					
			SIX6					
			SKIIP					
			SLC10A1					
			SLC25A21					
			SLC38A6					
			SLC39A9					
			SLC8A3					
			SNAPC1					
			SOCS4					
			SPG3A					
			SPTB					
			SPTLC2					
			SSTR1					
			STYX					

Table 3, continued

chr 6 (140–192cM)			chr 14 (35–75cM)			chr 11(130–164cM)		
Total	Function	Process	Total	Function	Process	Total	Function	Process
159 genes	56 genes	48 genes	198 genes	107 genes	69 genes	83 genes	12 genes	26 genes
			SYNJ2BP					
			SYT14L					
			TBPL2					
			TGFB3					
			THSD3					
			TIMM9					
			TMED8					
			TMP21					
			TRIM9					
			TTC6					
			TXNDC					
			VTI1B					
			WDHD1					
			WDR21A					
			WDR22					
			ZADH1					
			ZAP128					
			ZBTB1					
			ZBTB25					
			ZFP36L1					
			ZFYVE1					
			ZFYVE26					
			ZNF410					

Genes in **bold** are common to function and process GO terms, and * are genes that belong to the 24 'Alzgene database'.

overlapped. About 83 genes and nearly 8,321 HapMap Caucasian SNPs were recorded at the 11q22 QTL peak interval (116–161 cM) and our bioinformatics procedure reduced the candidate gene lists to 12 genes based on function and 26 based on process, with about 6 genes overlapping.

4. Discussion

Follow-up linkage analyses utilizing an additional 24 STRs with an average intermarker distance of ∼5 cM, confirmed suggestive linkages in the NIMH families on the CRIs located at 6q27, 14q22, and 11q25. The MLS scores for each CRI increased from the original scan and narrowing of the CRIs is also observed. Because of the likelihood of genetic heterogeneity in this complex disease, the heterogeneity LOD (HLOD) score statistic [31], which allows for linked and unlinked families in the sample, was performed on the original CIDR marker set and the follow-up set of markers in each region. In addition, the information content (IC), calculated in Genehunter Plus, for the CRI at 6q, 14q and 11q increased from a range of 0.30–0.54, 0.39–0.52, 0.32–0.52 in the original scan to 0.44–0.65, 0.45–0.62, 0.43–0.65 in the follow up, respectively. An IC estimate close to 0.7 is the theoretical maximum for sib pair families, which is the predominant structure in the NIMH families [32]. It is also worthwhile to mention that the chr. 14 'peak' is located at a good distance away from the most obvious candidate gene in the region, i.e. PSEN1. Hence, the above data and analyses suggest these regions may harbor additional loci impacting on risk to AD.

The reduced evidence for linkage at 3p26, initially identified as suggestive on the original genomic scan, may be attributed to confounding factors such as genetic or clinical heterogeneity of the disease, difficulty to detect genes of small effect, which reflects the difficulties and challenges that face the genetic mapping of complex traits like AD, or it may be a true false positive.

4.1. Classical gene discovery approach

In general, gene discovery for complex traits follows four strategic steps: whole genome linkage or association studies to identify chromosome candidate regions, fine-mapping by linkage or association studies of polymorphisms (Ex. SNPs) to identify the putative causal gene(s), sequence analysis of the pinpointed gene (s) to identify causal variant(s), and finally functional tests of the found variants [33]. However, it is becoming apparent that many, and perhaps most of the regions that

show linkage to a phenotype in multiple populations harbor more than one susceptibility locus [34]. For example, different asthma-related phenotypes are known to map to different locations within a broad linkage region [35]. Moreover, it is even likely that additional susceptibility loci reside within the same linkage peak and close to some of the positionally cloned genes, as has been shown for asthma [36]. Therefore, we propose that such regions are more likely to harbor multiple susceptibility loci, each with relatively small to moderate effects rather than large effects, on disease risk.

4.2. Limitations of the classical approach

The challenge in following-up linkage signals is to identify the genes that are responsible for the observed linkage results contained in such broad chromosomal regions (30 to 40 cM of recombination or 30–40 millions of DNA bases in length). Fine mapping of the linked region using more closely spaced STRs is limited by the low frequency of recombination events between any two closely spaced points in the genome and the limited number of highly informative STRs available, resulting in only marginal increases in IC.

If a CRI has been confirmed and narrowed with follow-up mapping, the region still may be relatively broad, perhaps 20–30 cM, such that potentially hundreds of genes may be contained within a single CRI. Traditionally, the 1-LOD-drop support interval flanking the peak MLS is used as a guide to define the critical region of an observed linkage peak [31], which could further narrow the region from 5–15 cM. However, a large number of genes may still be located in this narrowed region. Given current technology, an exhaustive analyses of all these genes is possible, however given current resources, comprehensive evaluation of all of the genes in a region of linkage is rarely possible, being both laborious and expensive. For example, the CRIs on chromosomes 6q27, 14q22, and 11q22 that we confirmed and narrowed to intervals of 52 cM, 40 cM, and 31 cM, respectively, still contain 159, 198, and 83 genes, respectively, according to the latest assembly (March 2006, NCBI build 36.1) of the human genome draft. Thus, a bioinformatics approach for identification and prioritization of candidate genes from the number of genes in confirmed CRIs remains a critical step following this classical approach. New approaches such as the current high-density genome-wide association (GWA) arrays also requires deep-sequencing, and if there are several SNPs in LD, then bioinformatics will help prioritize the genes to start deep sequencing on. GWA should be also considered as an initial step in the elucidation of susceptibility variants. There may be several regions that may harbor genetic variants that influence susceptibility to a specific disease. Several of the associated SNPs may fall within and near genes. Fine-mapping studies of these several regions are needed to confidently localize the signals to specific genes. Confirmation of the role of genes in the identified regions of interest will require prioritization of the genes and replication studies in other populations, and, ultimately, functional studies. More over, genome-wide association studies are promising, yet not always economically feasible or statistically desirable [37]. Therefore, one of the greatest challenges in disease association study design remains the intelligent selection of candidate genes.

4.3. Selection of candidate genes within CRI

There are potentially hundreds of potential candidate genes which could be investigated in association studies, family-based or case-control, using SNPs and programs such as LAMP [38] can be used to determine whether the entire linkage signal can be explained by a SNP. However, which candidate gene(s) to perform SNP genotyping for association testing is a conundrum. Hence, we must rely on the selection of positional candidate genes in the linkage region that have a known biological function related to our trait or are homologous to other genes in our phenotypic causal pathways. Combining mapping and arraying has been suggested as one approach to reduce the number of genes from QTL regions [39].

In the past few years, several groups have published bioinformatics methods for narrowing the lists of candidate genes (see the review by Tiffin et al. [34]. The fundamental assumption of the scheme for prioritizing candidate genes from linkage studies is that genes involved in or predisposing to a given polygenetic disease tend to share more commonalities in their molecular function, biological process or physiological pathway [40,41] than genes chosen at random or genes not involved in the same disease. Hence, Gene Ontology annotation terms will be enriched among genes linked to the trait [42] and such commonalities are often sufficient to identify these genes from regions containing hundreds of other genes.

Based on these assumptions, one way to narrow the list of candidate genes in AD is to search for annotation terms that are enriched among the systematic meta-analyzed AD candidate genes, that have been confirmed

by at least three case-control samples and cataloged in the "AlzGene database" as potential Alzheimer disease susceptibility genes (http://www.alzgene.org) relative to randomly sampled AD genes. Using these more biologically plausible GO function and process terms, we were able to reduce the number of positional candidate genes within each critical region (within the 1-LOD intervals) defined above. The list of genes can be initially prioritized based upon the number of GO terms assigned to each gene. Nevertheless, a large number of genes in the human genome are uncharacterized or poorly characterized, and the annotation of genes in Gene Ontology (GO) databases are incomplete and biased toward highly studied genes; thus, novel or poorly characterized genes could be missed. Moreover, GO is one approach and component to a more comprehensive systems biology approach. Additional sources of information such as gene expression, protein-protein interaction networks, tissue specificity, KEGG or BioCarta pathways, and sequence homology, could be integrated into a combined statistic, and biological confirmation of the candidate genes should be performed. Although the approach here still suffers from the bias introduced by pathways selected by biased researchers [30,43,44], this bias will be reduced once more genome wide association studies are incorporated into the AlzGene database, and gene annotation becomes more comprehensive. Finally, each investigator can then develop bioinformatics methods for further narrowing or prioritizing the list of candidate genes using a variety of selection variables such as expression profiles in neuronal tissue, the number of GO terms for each gene, evolutionary conservation, and patterns of gene duplication [33], and other systems biology approaches to determine which to interrogate first.

5. Conclusion

In 2003, we reported the results of a whole genome AD linkage study in the NIMH AD Genetics Initiative families [1]. In this study, we performed a follow up linkage study by genotyping 24 additional STR markers in the same sample of 437 families. The chromosome 6q27, 14q22, 11q27 and 3p26 CRIs were statistically analyzed under the same dominant inheritance model, resulting in confirmatory evidence for the CRIs at 6q27, 14q22, and 11q27. We were able to reduce the total number of genes in these regions to a list of plausible candidate genes using function and process GO terms derived from 24 meta-analyzed and confirmed AD genes on the Alzgene website. The approach, using bioinofrmatics tool and databases like Alzgene will facilitate in the identification and prioritization of candidate genes after the classical follow-up of promising linkage region and for deep-sequencing of several SNPs that are in LD in the current high-density genome-wide association array platforms.

6. Outlook of the experiment

Due to the evidence of linkage and the consistency of signals, we will be pursuing chromosome regions 6q27, 14q22, and 11q25 for possible candidate genes. We plan to first, undertake detailed fine mapping of these CRIs using dense SNPs selected from HapMap to narrow down the CRIs through linkage disequilibrium (LD) mapping. We will include in this phase the interrogation of candidate genes using tag SNPs from HapMap. We plan to priortize these genes based on the number of GO terms for each, and expression in neuronal tissue and determine which to interrogate first. Selection of SNPs in candidate genes will be based upon the location (coding/promoter vs. non-coding SNPs), type (if coding) (nonsynonymous vs. synonymous), as well as comparison with other species [45, 46] to identify highly conserved variants. Finally, sequencing within critical gene regions will need to be done in a systematic way to identify gene variants that may predispose to AD. These can then be confirmed in population-based studies, with further studies in animal models or other *in vivo* methods to ascertain its function in the disease process.

Acknowledgements

The authors would like to thank Micah Simmons for his technical support. The authors are extremely grateful to the families whose participation made this work possible. This work was supported by a grant from the NIMH (R01 NS045934-05).

References

[1] D. Blacker, L. Bertram, A.J. Saunders, T.J. Moscarillo, M.S. Albert, H. Wiener, R.T. Perry, J.S. Collins, L.E. Harrell, R.C. Go et al., Results of a high-resolution genome screen of 437 Alzheimer's disease families, *Hum Mol Genet* **12**(1) (2003), 23–32.

[2] R.T. Perry, H. Wiener, L.E. Harrell, D. Blacker, R.E. Tanzi, L. Bertram, S.S. Bassett and R.C. Go, Follow-up mapping supports the evidence for linkage in the candidate region at 9q22 in the NIMH Alzheimer's disease Genetics Initiative cohort, *Am J Med Genet B Neuropsychiatr Genet* **144B** (2007), 220–227.
[3] D.T. Selkoe, Alzheimer's disease: genes, proteins, and therapy, *Physiol Rev* **81** (2001), 741–766.
[4] R. Guttman, R.D. Altman and N.H. Nielsen, Alzheimer disease. Report of the council on scientific affairs, *Arch Fam Med* **8**(4) (1999), 347–353.
[5] A. Lobo, L.J. Launer, L. Fratiglioni, K. Andersen, A. Di Carlo, M.M. Breteler, J.R. Copeland, J.F. Dartigues, C. Jagger, J. Martinez-Lage, H. Soininen and A. Hofman, Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group, *Neurology* **54** (2000), S4–S9.
[6] R.F. Itzhaki and M.A. Wozniak, Herpes simplex virus type 1, apolipoprotein E, and cholesterol: a dangerous liaison in Alzheimer's disease and other disorders, *Prog Lipid Res* **1** (2006), 73–90.
[7] R.E. Tanzi, A genetic dichotomy model for the inheritance of Alzheimer's disease and common age-related disorders, *J Clin Invest* **104**(9) (1999), 1175–1179.
[8] R. Sherrington, E.I. Rogaev, Y. Liang et al., Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease, *Nature* **375** (1995), 754–760.
[9] E. Levy-Lahad, W. Wasco, P. Poorkaj et al., Candidate gene for the chromosome 1 familial Alzheimer's disease locus, *Science* **269** (1995), 973–977.
[10] A. M.Goate, M.-C. Chartier-Harlin, M.C., Mullan et al., Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease, *Nature* **349** (1991), 704–706.
[11] N. Ertekin-Taner, Genetics of Alzheimer's disease: a centennial review, *Neurol Clin* **25**(3) (2007), 611–667.
[12] A.D. Roses, A model for susceptibility polymorphisms for complex diseases: apolipoprotein E and Alzheimer disease, *Neurogenetics* **1**(1) (1997), 3–11.
[13] A.J. Slooter, M. Cruts, S. Kalmijn, A. Hofman, M.M. Breteler, C. Van Broeckhoven and C.M. van Duijn, Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study, *Arch Neurol* **55**(7) (1998), 964–968.
[14] M.L. Hamshere, P.A. Holmans, D. Avramopoulos, S.S. Bassett, D. Blacker, L. Bertram, H. Wiener, N. Rochberg, R.E. Tanzi, A. Myers, V.F. Wavrant-De, R. Go, D. Fallin, S. Lovestone, J. Hardy, A. Goate, M. O'donovan, J. Williams and M.J. Owen, Genome-wide linkage analysis of 723 affected relative pairs with late-onset Alzheimer's Disease, *Hum Mol Genet* (Aug 27, 2007), Epub ahead of print.
[15] L. Bertram and R.E. Tanzi, The current status of Alzheimer's disease genetics: what do we tell the patients? *Pharmacol Res* **50**(4) (2004), 385–396.
[16] A.S. Khachaturian, C.D. Corcoran; L.S Mayer, P.P Zandi and J.C.S Breitner, Apolipoprotein E 4 Count Affects Age at Onset of Alzheimer Disease, but Not Life time Susceptibility: The Cache County Study, *Arch Gen Psychiatry* **61** (2004), 518–524.
[17] M.I. Kamboh, Molecular genetics of late-onset Alzheimer's disease, *Ann Hum Genet* **68** (2004), 381–404.
[18] P. Kehoe, F. Wavrant-De Vrieze, R. Crook, W.S. Wu, P. Holmans, I. Fenton, G. Spurlock, N. Norton, H. Williams, N. Williams et al., A full genome scan for late onset Alzheimer's disease, *Hum Mol Genet* **8**(2) (1999), 237–245.
[19] M.A. Pericak-Vance, J. Grubber, L.R. Bailey, D. Hedges, S. West, L. Santoro, B. Kemmerer, J.L. Hall, A.M. Saunders, A.D. Roses et al., Identification of novel genes in late-onset Alzheimer's disease, *Exp Gerontol* **35**(9–10) (2000), 1343–1352.
[20] A. Myers, F. Wavrant De-Vrieze, P. Holmans, M. Hamshere, R. Crook, D. Compton, H. Marshall, D. Meyer, S. Shears, J. Booth et al., Full genome screen for Alzheimer disease: stage II analysis, *Am J Med Genet* **114**(2) (2002), 235–244.
[21] A. Sillen, C. Forsell, L. Lilius, K. Axelman, B.F. Bjork, P. Onkamo, J. Kere, B. Winblad and C. Graff, Genome scan on Swedish Alzheimer's disease families, *Mol Psychiatry* **11**(2) (2006), 182–186.
[22] L. Bertram, M.B. McQueen, K. Mullin, D. Blacker and R.E. Tanzi, Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database, *Nat Genet* **39**(1) (2007), 17–23.
[23] J.H. Lee, Mayeux, R. Mayo, D. Mo, J. Santana, V. Williamson, J. Flaquer, A. Ciappa, A. Rondon, H. Estevez et al., Fine mapping of 10q and 18q for familial Alzheimer's disease in Caribbean Hispanics, *Mol Psychiatry* **9** (2004), 1042–1051.
[24] G. McKhann, D. Drachman, M. Folstein et al., Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's disease, *Neurol* **34** (1984), 939–944.
[25] M.C. Tierney, R.H. Fisher, A.J. Lewis, M.L. Zorzitto, W.G. Snow, D.W. Reid and P. Nieuwstraten, The NINCDS-ADRDA Work Group criteria for the clinical diagnosis of probable Alzheimer's disease: a clinicopathologic study of 57 cases, *Neurology* **38**(3) (1988), 359–364.
[26] J.S. Collins, R.T. Perry, B. Watson, L.E. Harrell, R.T. Acton, D. Blacker, M.S. Albert, R.E. Tanzi, S.S. Bassett, M.G. McInnis, R.D. Campbell and R.C. Go, Association of a haplotype for tumor necrosis factor in siblings with late-onset Alzheimer disease: the NIMH Alzheimer Disease Genetics Initiative, *Am J Med Genet* **96** (2000), 823–830.
[27] A. Kong and N.J. Cox, Allele-sharing models: LOD scores and accurate linkage tests, *Am J Hum Genet* **61**(5) (1997), 1179–1188.
[28] L. Kruglyak, M.J. Daly, M.P. Reeve-Daly and E.L. Lander, Parametric and nonparametric linkage analysis: a unified multipoint approach, *Am J Hum Genet* **58** (1996), 1347–1363.
[29] S.A.G.E., Statistical analysis for genetic epidemiology. Version 5.0. (2002) Cork: Statistical Solutions Ltd.
[30] M.P. Mattson, Pathways towards and away from Alzheimer's disease, *Nature* **430**(7000) (2004), 631–639.
[31] J.D. Terwilliger and J. Ott, *Handbook of Human Genetic Linkage*, Baltimore: Johns Hopkins University Press, 1994, 320.
[32] D.M. Evans and L.R. Cardon, Guidelines for genotyping in genome-wide linkage studies: Single-nucleotide-polymorphism maps versus microsatellite maps, *Am J Hum Genet* **75**(4) (2004), 687–692.
[33] E. Lander and L. Kruglyak, Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results, *Nat Genet* **11**(3) (1995), 241–247.
[34] S. Hoffjan and C. Ober, Present status on the genetic studies of asthma, *Current Opinion in Immunology* **14** (2002), 709–717.
[35] B.A. Raby, E.K. Silverman, R. Lazarus, C. Lange, D.J. Kwiatkowski and S.T. Weiss, Chromosome 12q harbors multiple genetic loci related to asthma and asthma-related phenotypes, *Hum Mol Genet* **12** (2003), 1973–1979.

[36] E. Noguchi, Y. Yokouchi, J. Zhang, K. Shibuya, A. Shibuya and M. Bannai, Positional identification of an asthma susceptibility gene on human chromosome 5q33, *Am J Respir Crit Care Med* **172** (2005), 183–188.
[37] D. Thomas, Are we ready for genome-wide association studies? *Cancer Epidemiol Biomarkers Prev* **15** (2006), 595–598.
[38] M. Li, M. Boehnke and G. R. Abecasis, Joint modeling of linkage and association identifying SNPs responsible for a linkage signal, *Am J Hum Genet* **76**(6) (2005), 934–949.
[39] M.L. Wayne and L.M. McIntyre, Combining mapping and arraying: An approach to candidate gene identification, *Proc Natl Acad Sci USA* **99**(23) (2002), 14903–14906.
[40] N. Tiffin, E. Adie, F. Turner, H.G. Brunner et al., Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes, *Nucl Acids Res* **34** (2006), 3067–3081.
[41] T. Toyoda and Y. Takigawa, Selection of Candidate Genes for Polygenic Diseases by Utilizing Protein-Protein Interaction Networks, *Genome Informatics* **11** (2000), 286–288.
[42] F.S. Turner, D.R. Clutterbuck and C.A. Semple, POCUS: mining genomic sequence annotation to predict disease genes, *Genome Biology* (2003), 4:R75.
[43] M. Blurton-Jones and F.M. Laferla, Pathways by which Abeta facilitates tau pathology, *Curr Alzheimer Res* **3**(5) (2006), 437–448.
[44] M. Goedert and M.G. Spillantini, A century of Alzheimer's disease, *Science* **314**(5800) (2006), 777–781.
[45] C. Ferrer-Costa, M. Orozco and X. de la Cruz, Use of bioinformatics tools for the annotation of disease-associated mutations in animal models, *Proteins* **61** (2005), 878–887.
[46] K.M. Weiss, In search of human variation, *Genome Res* **8** (1988), 691–697.