

Universidade Federal de Pernambuco – UFPE
Centro de Ciências Biológicas – CCB
Mestrado em Ciências Biológicas

ROBERTA RODRIGUES DE LEMOS



Análise in silico de novos potenciais polimorfismos genéticos de risco na
Doença de Alzheimer em bancos de dados de *Microarrays*

RECIFE

2009

ROBERTA RODRIGUES DE LEMOS

Análise *in silico* de novos potenciais polimorfismos genéticos de risco na Doença de Alzheimer em bancos de dados de *Microarrays*

Dissertação apresentada ao programa de Mestrado em Ciências Biológicas da Universidade Federal de Pernambuco, como parte dos requisitos para obtenção do grau de mestre em Ciências Biológicas.

Orientador: João Ricardo Mendes de Oliveira

RECIFE
2009

Lemos, Roberta Rodrigues de

Análise in silico de novos potenciais polimorfismos genéticos de risco na Doença de Alzheimer em bancos de dados de Microarrays / Roberta Rodrigues de Lemos. – Recife: O Autor, 2009.

56 folhas: il., fig., tab.

Dissertação (mestrado) – Universidade Federal de Pernambuco. CCB. Departamento de Ciências Biológicas, 2009.

Inclui bibliografia.

1. Doença de Alzheimer 2. Bioinformática I Título.

**616.831
618.97**

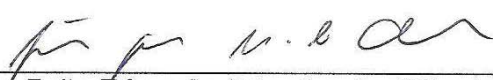
**CDU (2.ed.)
CDD (22.ed.)**

**UFPE
CCB – 2009- 038**

**Análise *in silico* de novos potenciais polimorfismos genéticos
de risco na Doença de Alzheimer em bancos de
dados de *Microarrays***

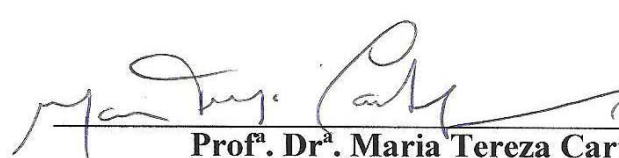
Roberta Rodrigues de Lemos

Banca Examinadora



Prof. Dr. João Ricardo Mendes de Oliveira

Orientador – UFPE / Dept. Neuropsiquiatria/LIKA



Prof.ª Dr.ª Maria Tereza Cartaxo Muniz

UPE / Dept. Genética UFPE



Prof.ª Dr.ª Cíntia Renata Rocha

UFPE / Dept. Bioquímica

À minha mãe Janette, minha grande
incentivadora e a toda minha família,
Dedico.

Agradecimentos

Agradeço a Deus por ter me conduzido ao lugar certo e por colocar no meu caminho pessoas maravilhosas.

Agradeço à minha mãe, Janette, minha grande e fiel incentivadora. Obrigada mãe, por sempre estar presente nos momentos alegres e tristes da minha vida e por ter me ensinado a agradecer por tudo o que nos acontece.

Obrigada Família, por compartilhar comigo minhas conquistas e pela grande torcida e compreensão.

Agradeço ao professor João Ricardo, por ter acreditado e pelo seu contínuo exemplo de paciência, sabedoria, alegria e perseverança.

Agradeço aos meus queridos amigos que com sua alegria e descontração fizeram dos meus dias de trabalhos e estudos, dias felizes.

Agradeço aos programas CAPES, FACEPE, PROPESQ-UFPE, pelo apoio financeiro.

Sumário

Lista das Abreviaturas

Lista das Figuras

Lista das Tabelas

Resumo

Abstract

1. Revisão da Literatura	1
1.1 Neurodegeneração e Genética	1
1.2 <i>Single Nucleotide Polymorphisms</i> – SNPs	2
1.3 A genética da Doença de Alzheimer	4
1.4 Polimorfismos genéticos na DA	7
1.5 <i>Sequence Expressed Tags</i> – ESTs	8
1.6 Estudos de <i>Microarrays</i> de expressão e genotipagem	9
1.7 Análise <i>in silico</i> de SNPs a partir de ESTs	11
1.8 Bioinformática: <i>Basic Local Alignment Search Tool</i> - BLAST x Smith-Waterman	13
2. Justificativas	14
3. Objetivos	15
3.1 Gerais	15
3.2 Específicos	15
4. Capítulo I	16
Linkage studies in familial idiopathic basal ganglia calcification: Separating the wheat from the chaff. <u>Am J Med Genet B Neuropsychiatr Genet.</u> 2008 Jul 28. In Press. DOI 10.1002/ajmg.b.30839	
5. Capítulo II	17
The Genetics of Alzheimer's Disease in Brazil: 10 Years of Analysis in a Unique Population. <u>J Mol Neurosci.</u> v. 37, 74-79, 2009	
6. Capítulo III	18
In silico identification of new genetic variations as potential risk factors for Alzheimer's Disease in a Microarray oriented simulation. <u>J Mol Neurosci.</u> 2009. In Press DOI: 10.1007/s12031-009-9191-x.	
7. Anexo	19
Resultados complementares	
Identificação <i>in silico</i> de Deleções e Transições	22

Deleções: maiores classes de mudanças detectadas	22
Mutações sinônimas e não Sinônimas	23
Tradução de seqüências: <i>Frameshift</i> e <i>Stop codons</i>	25
Variações na regiões UTRs	26
Análise do percentual GC	27
Genes <i>up-regulated</i> e <i>Down-regulated</i> envolvidos com a via inflamatória e degradação das placas <i>B</i> -amilóide	29
8. Conclusão	31
9. Referências Bibliográficas	48

Lista de Abreviaturas

APOE – Apolipoproteína E

APP - Proteína Precursora Amilóide

BLAST- Basic Local Alignment Search Tool

cDNA- Complementar Desoxyribonucleic acid

DA - Doença de Alzheimer

DF- Doença de Fahr

DN - Degeneração Neuronal

DNA - Desoxyribonucleic acid

dbNCBI- Database National Center for Biotechnology Information

EOAD- Early Onset Alzheimer's Disease

ESTs - Express Sequence Tags

INDELS – Inserções e Deleções

LOAD- Late Onset Alzheimer's Disease

MAPs- Proteínas Associadas ao Microtubulo

NCBI - National Center for Biotechnology Information

PCR - Reação em cadeia da polimerase

PSEN 1 - Presenilina 1

PSEN 2 - Presenilina 2

RNA- Rybonucleic acid

RNAm- Rybonucleic acid messenger

SAP- Single Amino acid Polymorphism Effect Prediction

SNPs - Single Nucleotide Polymorphisms

UCSC – University of California Santa Cruz

UTR – Untranslated region

Lista das Figuras

Figura 1. O programa SAP classifica a alteração protéica, originada pelos SNP, em uma mutação associada à doença ou mutação neutra, usando apenas a informação derivada da sequência protéica.	33
Figura 2. O programa miRBASE é uma ferramenta de alinhamento de trechos de seqüências de RNA contra um repositório de seqüências de microRNA	34
Figura 3. Identificação dos códigos de ESTs e das posições na região UTR e CDs, a parte sublinhada representa variações previamente identificadas nos genes CTSC isoforma A durante o processo de validação virtual.	35
Figura 4. Alinhamento de duas ESTs (BM926451) e (BG024434) com o exon 5a e o o miRNA (has-miR-939). A seta indica a posição da deleção.	35
Figura 5. Análise do conteúdo GC dos exons 2, 4 e 5a do gene CD83 isoforma a.	36
Figura 6. Análise do conteúdo GC dos exons 3, 5 e 7 e 9 do gene COX2	36
Figura 7. Análise do conteúdo GC dos exons 6 e 7 do gene IL-1 α .	37
Figura 8. Análise do conteúdo GC dos exons 6 e 7 do gene IL-1 β	37
Figura 9. Análise do conteúdo GC dos exons 1,2,15, e 17 do gene ECE isoforma A.	38
Figura 10. Análise do conteúdo GC dos exon 3 do gene ECE isoforma C.	38
Figura 11. Análise do conteúdo GC dos exon 2b do gene CXCR4 isoforma B.	38
Figura 12. Análise do conteúdo GC dos exons 2,3,4,5,7,8,9, 10, 11 do gene TNFAIP2.	39
Figura 13. Análise do conteúdo GC dos exons 7,11,14,15,16 do gene PLA2G4A.	40
Figura 14. Análise do conteúdo GC dos exons 6, 7, 9 do gene CTSC.	41

Lista de Tabelas

Tabela 1. Resumo dos estudos de associação caso-controle na população Brasileira.	8
Tabela 2. Lista dos nomes, localização e códigos de acesso dos quatro genes estudados	21
Tabela 3. Resumo da análise <i>in silico</i> dos genes: TNFAIP2, ECE2 isoforma A e C, CTSC isoforma A, CXCR4 isoforma B.	22
Tabela 4. Deleções preditas que causam prematuras proteínas truncadas.	42
Tabela 5. Adicionais polimorfismo no gene TNFAIP2.	43
Tabela 6. Adicionais polimorfismo no gene CTSC.	44
Tabela 7. Adicionais polimorfismo no gene ECE 2 Isoforma A e C.	46
Tabela 8. Adicionais polimorfismo no gene CXCR4 isoforma B.	47

Resumo

Estudos de genômica e proteômica sobre fatores de risco associados à desordens neurodegenerativas requerem urgentemente proposições de abordagens complementares para integrar a grande quantidade de dados gerados. Neste trabalho foi proposto uma metodologia que envolve ferramentas de Bioinformática, no qual estudos de *microarrays* de expressão foram usados como a única origem de genes candidatos, para revelar variações novas, a partir de seqüências públicas de Expressed site tags (ESTs). Foram selecionados nove genes, sete de um estudo de tecido do hipocampo área Cornu Ammonis (CA1) e dois do tecido do lóbulo parietal inferior (IPL) ambos de pacientes com Doença de Alzhiemer (DA), a maioria desses genes está envolvido com o sistema imune, escolhidos neste trabalho por fazer parte de um das linhas de pesquisa do grupo. O CLCbio Workbench Combined[®] versão 3.6.2. foi inicialmente usado para construção do banco de dados de ESTs e recuperação dos arquivos de RNAm respectivamente a partir do Golden path of University of California Santa Cruz (UCSC) e National Center for Biotechnology Information (NCBI), na etapa seguinte foram realizados múltiplos alinhamentos e o algoritmo usado foi o Smith-Waterman. Um total de 479 seqüências de ESTs foram selecionadas depois da aplicação de parâmetros apropriados de estringência utilizados para minimizar erros de alinhamentos. A anotação revelou várias classes de variações, a maioria delas sendo deleções (415), mas também foram observadas transições (253), transversões (52), mutações sinônimas (48), não sinônimas (400) e Single Nucleotide Polymorphisms (SNPs) em Untranslated Regiões (UTRs) (50). Deleções são frequentemente associadas as principais síndromes genéticas com características dismórficas. No entanto, vários recentes trabalhos mostram que microdeleções comuns podem ser associadas com desordens neuropsiquiátricas como esquizofrenia, autismo e retardo mental, detectas em várias etnias através de experimentos de sequenciamento. A validação virtual confirmou que algumas dessas variações identificadas foram previamente anotadas e confirmadas em amostras de DNA, demonstrando que esse método é uma maneira viável para detectar variações genéticas que merecem ser exploradas futuramente em estudos de associação de fatores genéticos de risco para a DA.

Palavras chave: In silico, Doença de Alzheimer, Bioinformática, ESTs, Microarrays.

Abstract

Genomic and Proteomic studies of neurodegenerative disorders require complementary approaches to integrate the massive amount of data generated in high throughput experimental procedures. We propose a Bioinformatics pipeline in which expression studies guide the selection of candidate genes that should be screened for potential new genetic variations from a public Expressed site tags (ESTs) database. Motivated by the former interest of our group in genetic polymorphisms involved with the immune system, we selected nine genes from a previous expression microarrays study of hippocampal Cornu Ammonis (CA1) area and Inferior Parietal Lobe (IPL) of Alzheimer's Disease subjects (AD). The CLCbio Workbench Combined[®] version 3.6.2. was initially used to build ESTs and mRNA files retrieved respectively from the Goldenpath of University of California Santa Cruz (UCSC) and National Center for Biotechnology Information (NCBI) databases and latter to perform multiple batches of Smith-Waterman alignments. A total of 479 ESTs sequences were selected after proper stringent parameters were applied to the first set of mismatches. The annotation revealed various classes of variations, most of them deletions (415), but also transitions (253), transversions (52), synonymous (48), non synonymous (400) and Single Nucleotide Polymorphisms (SNPs) in Untranslated Regions (UTRs) (50). Deletions are often associated to major genetics syndromes with dysmorphic features. However, various recent studies show that common microdeletions might be highly associated with common neuropsychiatric disorders such as schizophrenia, autism, mental retardation or even in various ethnicities detected in whole genome sequencing experiments. A virtual validation confirmed that some of the variations identified were previously reported and confirmed in DNA samples, showing that this method is a feasible way to detect genetic variations that merit further exploration in AD genetic risk factor association studies.

Key words: In silico, Alzheimer's disease, Bioinformatics, ESTs, Microarrays.

1. Revisão da Literatura

1.1. Neurodegeneração e Genética.

A degeneração neuronal (DN) é um fenômeno complexo, com etiologia multifatorial e presente em várias doenças neuropsiquiátricas, tais como: Doença de Alzheimer (DA) e Doença de Fahr (DF). Este fenômeno é evidenciado por sinais e/ou sintomas progressivos, com um amplo espectro de início, provocando lesões irreversíveis e que já foram bem caracterizadas em análises anatomopatológicas e radiológicas (GESCHWIND *et al.*, 1999; WINKLHOFFER *et al.*, 2008).

Pouco se conhece sobre os fatores de risco genéticos para as doenças neurodegenerativas e frequentemente os estudos prévios neste tema estão sendo revistos e analisados de uma maneira crítica no sentido de aumentar a confiabilidade dos achados já existentes e de novos dados.

Há mais de dez anos nosso grupo vem investigando alguns destes transtornos, do ponto de vista molecular e genético. Para alguns temas, como a DA e a DF, os estudos procuram esclarecer o cenário genotípico e fenotípico destes tipos de doenças (OLIVEIRA *et al.*, 2007; 2009; OLIVEIRA; LEMOS, 2008).

Estudos de genômica e proteômica sobre fatores de risco associados à DN requerem urgentemente proposições de abordagens complementares para integrar a grande quantidade de dados gerados (MILLER *et al.*, 2008).

A disponibilidade de novas ferramentas de Bioinformática associadas aos estudos de genotipagem na população está permitindo analisar e sintetizar de maneira cada vez mais conclusiva e definitiva o extenso conjunto de dados referentes aos fatores de risco genético para doenças neuropsiquiátricas tais como a DA.

1.2. *Single Nucleotide Polymorphisms* – SNPs.

SNPs são uma das mais comuns fontes de variações genéticas (NICKERSON *et al.*, 1998; WANG *et al.*, 1998). Essas variações podem ser mudanças de base classificadas de transições ou transversões, *indels* inserções ou deleções de bases. Os SNPs podem ser localizados em regiões codificadoras (cSNPs) ou não codificadoras (nSNPs), distribuídos aleatoriamente ou formando *clusters* em alguns genes (NICKERSON *et al.*, 1998).

As variações que mudam o quadro de leitura podem causar alterações relevantes no funcionamento normal do organismo, como o que acontece com o alelo variante $\epsilon 4$ do gene APOE (WEISS & TERWILLIGER, 2000).

Os estudos de SNPs têm muitas aplicações na caracterização da estrutura dos genes, na história de grupos populacionais e na identificação de loci responsáveis pela variação fenotípica de origem multifatorial, incluindo doenças de susceptibilidade. Apesar de terem surgido por mutações, muitas das posições que contêm SNPs apresentam baixas taxas de mutações, e podem ser utilizadas como marcadores moleculares estáveis no mapeamento de genes (GIBSON; MUSE, 2004; LESK A, 2008).

Um projeto internacional denominado “HapMap” é uma parceria de cientistas de diversos países como: Canadá, China, Japão, Nigéria, Reino Unido e Estados Unidos, que tem como objetivo descrever padrões de variações de seqüências do DNA humano.

Cerca de 10 milhões SNPs existem na população humana, quando o projeto iniciou 2.8 milhões de SNPs foram provenientes de bancos de dados públicos como *data base* SNP- dbSNP do National Center Biotechnology Information - NCBI, em setembro de 2003 outros 2.8 milhões já foram descritos e o projeto continua descrevendo novos SNPs (www.hapmap.org).

A grande quantidade de dados sobre SNPs está disponível em bancos de dados como o: TSC, HGVBBase, JSNP e dbSNP (HIRAKAWA *et al.*, 2002; THORISSON; STEIN, 2003; FREDMAN *et al.*, 2004; WHEELER *et al.*, 2005).

Embora muitos dados a cerca dos genes e variações genéticas associadas a doenças estejam no domínio público encontrar esses achados não é uma tarefa trivial. Nesse sentido, bancos de dados se aperfeiçoaram para disponibilizar de maneira cada vez mais pratica esse conjunto de informações é o caso da ferramenta Gene Prospector, indexada ao site HuGE Navigator, onde há uma integração de conhecimento básico de associação genética e epidemiologia do genoma humano (YU W *et al.*, 2008; YU WEI *et al.*, 2008).

Outros bancos de dados disponíveis gratuitamente são mais refinados na busca de informação biológica associada ao SNP, se especializando em inferir significado tanto para mutações identificadas em regiões codificantes, que alteram a sequência protéica, quanto para variações localizadas nas 3' *untranslated regions* - 3'UTRs.

O programa *Single Amino acid Polymorphism Effect Prediction* (SAP) classifica a alteração protéica, originada pelo SNP, em uma mutação associada à doença ou mutação neutra, gerando índices finais resultantes para a versão selvagem e para a versão mutada (HU; YAN, 2008).

Estes índices são calculados através do AAindex (*Amino Acid Index Database*) que armazena propriedades físico-químicas e bioquímicas de aminoácidos e de pares de aminoácidos (KAWASHIMA *et al.*, 1999).

MicroRNA (miRNA) são pequenas moléculas (~21 nucleotídeos) não codificantes de RNAm, originadas de transcritos de RNAm ou através de processo de *splicing* alternativo, neste último caso sendo precursores de sequências curtas intrônicas denominadas *mirtrons* (BEREZIKOV *et al.*, 2007).

Essas miRNAs funcionam como sítios de complementaridade que alinham com a sequência transcrita de RNAm na região 3'UTR e dessa maneira regula a expressão, através da repressão da tradução ou degradação do transcrito, alguns trabalhos associam variações genéticas nas miRNA à doenças. (CHUZHANOVA *et al.*, 2007; GRIFFITHS-JONES *et al.*, 2007; LUI *et al.*, 2007).

A ferramenta miRBASE é um repositório de 6.397 seqüências de miRNA, onde 678 são de origem humana, nesse site é possível realizar alinhamentos entre seqüências de RNAm contra seqüências de miRNA.

1.3. A genética da Doença de Alzheimer

A DA é a principal forma de demência em idosos, cuja incidência estimada para 2025 é de 34 milhões de pacientes em todo o mundo (WHO, 2006). O número estimado de pacientes Brasileiros com essa doença é de 1 milhão e 200 mil (www.alzheimermed.com.br).

A DA é uma desordem neuropsiquiátrica degenerativa, caracterizada inicialmente por déficits de memória recente e memória a longo prazo, eventualmente progredindo até que o paciente apresente um comprometimento significativo da linguagem e da capacidade de desempenhar tarefas. Alterações comportamentais como psicose, alterações do humor, do sono, agitação psicomotora e agressividade, estão presentes ao longo da doença (FORLENZA, 2005; PASTORINO *et al.*, 2006; DUBOIS *et al.*, 2007).

Do ponto de vista macroscópico, o cérebro do paciente com DA revela diminuição do volume, atrofia na região temporal, parietal e do córtex frontal, reflexos diretos da morte celular neuronal e estando estas regiões associadas com o processo de formação e evocação de memória (MATTSON, 2004).

A forma precoce (*Early Onset Alzheimer's Disease* - EOAD) se inicia antes dos 65 anos, tem herança autossômica dominante e é identificada em grupos familiares específicos, correspondendo a aproximadamente 10% dos casos. Até o momento, várias mutações foram verificadas em três genes, sendo eles: proteína precursora amilóide (APP), presenilina-1 (PSEN-1) e presenilina-2 (PSEN-2), localizados nos cromossomos 21, 14 e 1 respectivamente (HEESE *et al.*, 2006; PASTORINO *et al.*, 2006; DUBOIS *et al.*, 2007).

Postula-se que durante o processamento da proteína amilóide ocorra uma clivagem proteolítica pela ação de três isoenzimas secretases (α , β e γ), dando origem a fragmentos protéicos insolúveis ($A\beta_{1-40}$ e $A\beta_{1-42}$). Estes resíduos depositados na região extracelular formam as placas senis (HAASS *et al.*, 1992). As presenilinas também estariam de alguma forma envolvidas nesse processamento dos resíduos e alguns autores sugerem que mutações nestes genes agiriam como secretases durante o processo proteolítico (XIA W, 2001).

A forma tardia (*Late Onset Alzheimer's Disease* - LOAD) ocorre após os 65 anos e corresponde à maioria dos relatos em DA, tendo como modelo a herança multifatorial com contribuição de fatores genéticos e ambientais. O principal gene relacionado com essa forma é o alelo $\epsilon 4$ da Apolipoproteína E (APOE), localizado no cromossomo 19 (HIRSCH-REINSHAGEN *et al.*, 2008).

O alelo $\epsilon 4$ da APOE é o mais importante fator de risco para LOAD, modificando a evolução clínica do quadro demencial, a resposta a agentes anti-colinesterásicos e modulando também modificações anatômicas e metabólicas detectáveis através de exames de neuroimagem (OLIVEIRA *et al.*, 1999; POIRIER, 1999; SMALL *et al.*, 2000; BORRONI *et al.*, 2002).

O gene APOE produz três isoformas distintas: $\epsilon 2$ $\epsilon 3$ $\epsilon 4$, sendo que a isoforma $\epsilon 2$ agiria como fator protetor contra a doença, enquanto a isoforma $\epsilon 4$ seria um fator de predisposição. Acredita-se que a isoforma $\epsilon 4$ estaria contribuindo para a formação das placas, no sentido de que quando presente no meio extracelular neuronal ela é recaptada preferencialmente pelo polipeptídeo ($A\beta_{40-42}$), permitindo um maior tempo de exposição desse agregado e desencadeando a precipitação e a formação das placas senis (ST-GEORGE-HYSLOP, 2000; HIRSCH-REINSHAGEN *et al.*, 2008).

No entanto, essa variante genética está longe de explicar todo o processo neurodegenerativo e vários outros polimorfismos têm sido associados à essa devastadora doença.

Um outro achado macromolecular associado à DA são os emaranhados neurofibrilares, que são constituintes intracelulares e se formam pela hiperfosforilação da proteína TAU.

A proteína TAU é expressa nos axônios e faz parte da classe das proteínas associadas ao microtúbulo (MAPs), sendo estas responsáveis pela polimerização, estabilização e propriedade dinâmica dos microtubulos. As MAPs são polímeros de dímeros de α e β tubulina que mediam muitas funções nos neurônios, incluindo transporte de organelas, estabelecimento da forma celular, manutenção e alongação do axônio (DEHMELT L; HALPAIN S, 2004).

Os depósitos de β -amilóide e os emaranhados neurofibrilares podem estar presentes em outras doenças neurodegenerativas que se apresentam com déficits cognitivos ou neurológicos (LEE *et al.*, 2001).

Alterações inflamatórias são observadas nos cérebros de pacientes com DA, principalmente a cerca das placas amilóides onde a atividade das microglias é intensa. Uma vez estimulada, a microglia desencadeia uma série de mediadores pró-

inflamatórios, tais como: citocinas, componentes complementares, vários radicais livres e o óxido nítrico. Todos estes potencialmente contribuem para a disfunção neuronal e eventualmente morte celular (GRIFFIN, W. S *et al.*, 1998; GRIFFIN, W. S. T, 2000; HENEKA; O' BANION, 2007).

1.4. Polimorfismos genéticos na DA.

Acredita-se que os SNPs sejam as mais freqüentes fontes de mutações que predispõem à LOAD. A literatura nos revela que existem cerca de 4 milhões de SNPs já identificados no genoma humano e prevê ainda outras seqüências a serem identificadas para que se possa ter uma ampla associação com doenças do Sistema Nervoso Central (FREUDENBERG-HUA *et al.*, 2003).

Recentemente, um grande acervo de dados foi produzido a cerca dos fatores genéticos de risco para a DA, muitos desses provenientes de estudos de caso-controle, estando essas informações disponibilizadas em um banco de dados público nomeado de ALZGENE, acessível pelo site (www.alzforum.org).

Mais de 300 polimorfismos genéticos estão envolvidos com a DA, identificados em diferentes etnias ao redor do mundo. No entanto, somente alguns desses foram confirmados independentemente, demonstrando que essa condição é poligênica e com um padrão de herança genético complexo (BERTRAN *et al.*, 2007).

Oliveira *et al.* (2009) reuniram resultados sobre genes e marcadores genéticos envolvidos na DA, publicados desde 1997 pelo próprio grupo, além de outros grupos brasileiros. Na tabela 1 estão sumarizados os dados dos estudos do tipo caso controle acerca dos seguintes marcadores ou genes: Apolipoproteína E (APOE), 5-Hidroxytryptamine transporter long polymorphic region (5HTTLPR), Brain-derived neurotrophin factor (BDNF) Monoamine Oxidase A (MAO-A), Interleucina (IL) -1 α e

IL-1 β , dois *Simple Sequence Tandem Repeats Polymorphisms* (DXS1047 e D10S1423) e o Methylenetetrahydrofolate reductase enzyme (MTHFR).

Nesse mesmo trabalho foram apresentados resultados sobre estudos de Ligação a respeito dos seguintes marcadores em 4 famílias com DA: D1S158, D1S479, F13B, D1S53, D1S179, D1S217, D1S249, D1S175, D1S178, D14S52, D14S53, D14S42, N141I, D21S120, D21S13E, APP, APOCII (MFD5), D19S278, D19S178, D19S224, D19S47 e D19S75, nenhuma dessas, mostraram um padrão distinto e significativo de segregação (OLIVEIRA *et al.*, 2009).

Tabela1. Resumo dos estudos de associação caso-controle na população Brasileira

Referência	Polimorfismos	Nível de significância	Tamanho da amostra
Oliveira et al. (1997)	ApoE- ϵ 4	$P < 0,05$	57
Oliveira et al. (1998)	5-HTTLPR	$P < 0,05$	81
Nishimura et al. (2000)	DXS1047	$P > 0,05$	130
Nishimura et al. (2001)	D10S1423	$P > 0,05$	130
Nishimura et al. (2004)	BDNF (C-270T)	$P > 0,05$	188
	BDNF (C-270T) + ApoE- ϵ 4	$P < 0,05$	
	MAOA	$P > 0,05$	
Nishimura et al. (2005)	MAOA + ApoE- ϵ 4+		128
	5-HTTLPR	$P < 0,05$	
	IL-1 α and IL-1 β	$P > 0,05$	
Dados não publicados	ApoE- ϵ 4	$P < 0,05$	195
Fernandez et al. (2005)	MTHFR (C-677T)	$P > 0,05$	30
	ApoE- ϵ 4	$P < 0,05$	
da Silva et al. (2006)	MTHFR (C-677T)/(A-1298C	$P > 0,05$	43

1.5. *Sequence Expressed Tags* – ESTs.

Estudos de expressão constituem uma rica fonte de *Sequence Expressed Tags* ESTs, que são seqüências terminais curtas de DNA complementar (cDNA) que descreve padrões de transcrição dos genes, estas podem ser utilizadas como “matéria prima” para a localização de variações no genoma (ADAMS *et al.*, 1991)

O sequenciamento do genoma humano tem gerado importantes informações sobre o funcionamento básico do DNA e aumentado o interesse na pesquisas que comparam diferenças genéticas entre indivíduos (LANDER *et al.*, 2001).

Medições dos níveis de RNAm podem ser armazenados em bancos de dados de expressão, por meio das ESTs, esses bancos em sua maioria, contêm campos que indicam os tecidos de origem e/ou a localização do tecido subcelular, o estágio de desenvolvimento, as condições de crescimento, e as quantificações dos níveis de expressão, dentro do GenBank o dbEST contém atualmente quase 43 milhões de entradas de 1.200 espécies, as seqüências humanas lideram o ranking com 7.974.440 entradas (LESK A, 2008).

Nos bancos de dados públicos como NCBI e Golden Path da Universidade da Califórnia, Santa Cruz (UCSC), essas ESTs ficam disponíveis para vários tipos de pesquisas.

Um número crescente de seqüências disponíveis para análise de Bioinformática já tem sido explorada com várias técnicas de análises de dados e softwares e têm um potencial ilimitado de investigação se combinados e guiados por estudos funcionais e de expressão (PICOULT-NEWBERG *et al.*, 1999; USECHE *et al.*, 2001; FORMENT *et al.*, 2008).

1.6. Estudos de *Microarrays* de expressão e genotipagem.

Os estudos com *microarrays* permitem a medição em paralelo da expressão de milhares de genes em um único experimento. O princípio básico da tecnologia consiste da hibridização de RNA mensageiro, extraído de células de uma amostra de tecido de interesse, com o cDNA representativo de uma numerosa classe de genes.

O cDNA é imobilizado e distribuído em forma matricial sobre uma superfície apropriada. Cada gene é representado por um ou mais “spots” de cDNA na superfície da

microarray e este esquema de endereçamento espacial é o que permite a investigação em paralelo de milhares de genes em um único experimento.

O RNA mensageiro é previamente marcado com uma substância fluorescente, o que permite, após a hibridização e lavagem do excesso de material, digitalizar uma imagem da *microarray*. Os spots que se apresentam luminosos correspondem a genes que estão sendo expressos na amostra (GESCHWIND *et al.*, 2000).

A análise de dados de *microarrays* consiste no uso de métodos gráficos, tais como agrupamento e redução de dimensionalidade e de técnicas estatísticas, tais como testes de hipóteses e projeto de classificadores (RENCHE, 2002; BRAGA-NETO; DOUGHERTY, 2005).

Algoritmos de seleção de atributos podem encontrar automaticamente genes que melhor discriminem entre as classes presentes na amostra (KIM *et al.*, 2002); classificadores projetados usando esses genes como variáveis podem servir como ferramentas de diagnóstico/prognóstico clínico (VAN DE VIJVER *et al.*, 2001). A eficácia de tais classificadores é avaliada através de métodos robustos de estimação de erro, que indicam o grau de discriminação atingida pelo conjunto de genes considerado (BRAGA-NETO; DOUGHERTY, 2004).

Além disso, os genes discriminantes e classificadores assim descobertos podem levar a um entendimento básico dos mecanismos moleculares envolvidos. Desta forma, hipóteses científicas são automaticamente geradas, a partir do massivo conjunto de dados de expressão gênica obtidos com as *microarrays*. É importante ressaltar que essas hipóteses devem ser validadas por meio de experimentos bioquímicos tradicionais conduzidos independentemente.

Uma outra abordagem semelhante consiste em *microarrays* para genotipagem, que permitem detectar a presença de milhares de SNPs. Estas *microarrays* de SNPs

usam o mesmo princípio básico das *microarrays* de DNA, com a diferença que em vez do uso de RNA mensageiro para hibridização é utilizado DNA obtido dos pacientes e amplificado por *Polymerase chain reaction* – PCR (HACIA, 1999).

As sondas de DNA na *microarray* de genotipagem são organizadas em grupos de cinco, que representam os quatro valores possíveis de uma base em uma determinada posição e uma quinta variante que representa deleção. Esta abordagem vem sendo usada na detecção de padrões de SNPs em grande escala em estudos populacionais ou de associação com amostras de DNA humano.

Bancos de dados de *arrays* de expressão podem ser acessados gratuitamente, os principais são: o do European Bioinformatics Institute-EBI (www.ebi.ac.uk/arrayexpress), do NCBI (www.ncbi.nlm.nih.gov/geo) e Stanford Microarray Database–SMD (www.stanford.edu/microarray/smd).

1.7. Análise *in silico* de SNPs a partir de ESTs

Diante do grande número de genes de risco para a DA que já foram identificados, além dos milhares de candidatos, os estudos agora se concentram em entender como esses genes interagem entre si, para controlar o surgimento e progressão da doença. A metodologia das *microarrays* vem se mostrando cada vez mais promissora para atingir este objetivo (COLANGELO *et al.*, 2002; WEERARATNA *et al.*, 2007; MILLER *et al.*, 2008).

Resultados de experimentos de larga escala tais como, *microarrays* de expressão e de genotipagem relacionados a uma doença, podem ser confrontados com estudos de identificação e análise de variações de nucleotídeos, através do uso combinados de métodos analíticos seriados que já estão sendo otimizados (COPPOLA *et al.*, 2008).

Ferramentas de Bioinformática foram construídas com diferentes propostas e são atualmente curadas e melhoradas ao mesmo tempo para processar informações a respeito

das ESTs, permitindo o alinhamento e a comparação de numerosas quantidades de seqüências com o objetivo de identificar variações genéticas como os SNPs (USECHE *et al.*, 2001; BARKER *et al.*, 2003; FORMENT *et al.*, 2008; NAVRATIL *et al.*, 2008; TANG *et al.*, 2006, 2008).

Alguns trabalhos já estudaram amostras de pacientes com DA através de *microarrays* e são uma fonte importante de genes candidatos, para o estudo de novos fatores de risco. (COLANGELO *et al.*, 2002; EMILSSON *et al.*, 2006; BROOKS *et al.*, 2007; WEERARATNA *et al.*, 2007).

Uma integração de estudo de expressão e genotipagem foi realizado por Coppola *et al.* (2008) que identificou variações do tipo IVS7-1G>C e c.675_676delCA no gene da Progranulina em duas amostras de DNA de dois indivíduos com notáveis níveis de expressão diminuída para esse gene em comparação com prévios resultados de micorarrays de expressão nesse mesmo estudo.

Pretendendo trabalhar nessa linha nosso grupo propôs uma metodologia que envolve ferramentas de Bioinformática, onde prévios estudos de *microarrays* de expressão de DA, foram usados como a única origem de genes candidatos, que foram minuciosamente examinados em bancos de dados públicos, para revelar variações novas candidatas, que possam ser depois exploradas como potenciais fatores de risco de susceptibilidade.

Neste estudo, identificamos deleções que variaram de 1 a 10pb nas regiões exônicas de um grupo de genes que faz parte da via inflamatória e que apresentaram níveis de expressões alterados em amostras de tecido da área *Cornu Ammonis 1* (CA1) sub-região do hipocampo. (LEMOS *et al.*, 2009).

Este tipo de iniciativa é a chave para maximizar a nossa compreensão dos subjacentes painéis moleculares dessa doença.

1.8. Bioinformática: *Basic Local Alignment Search Tool* - BLAST x Smith-Waterman.

O alinhamento de seqüências biológicas é uma forma rápida de agregar informações sobre uma seqüência desconhecida e compará-la com um banco de dados de seqüências de funções conhecidas.

Esta comparação é feita através da sobreposição par a par entre as seqüências. Isto é, se o banco de dados possuir 1000 seqüências conhecidas serão realizados 1000 alinhamentos com a seqüência de entrada.

Os algoritmos mais comuns usados para pesquisa de similaridade em bancos de dados são: Smith-Waterman (SMITH-WATERMAN, 1981) e o BLAST (ALTSCHUL *et al.*, 1990; TATUSOVA; MADDEN, 1999).

O algoritmo Smith-Waterman é o método que aperfeiçoa o alinhamento local, pois leva em consideração a presença de *indels* e permite a comparação de seqüências de tamanho diferentes, bem como alinhamentos com apenas alguns trechos conservados (SMITH-WATERMAN, 1981). Porém, o tempo de execução é bem maior em relação ao BLAST.

O BLAST é um algoritmo que foi desenvolvido a partir do Smith-Waterman, com o objetivo de melhorar o tempo de execução. No entanto, este método não tem a mesma precisão que o Smith-Waterman para seqüências pequenas (ALTSCHUL *et al.*, 1990).

Para os estudos biológicos é essencial que esses alinhamentos indiquem uma real similaridade entre as seqüências analisadas. O índice de similaridade considera a probabilidade de o alinhamento ter ocorrido ao acaso. Os valores do *Score* e da identidade quantificam o grau de similaridade entre as seqüências comparadas.

Outro índice medido é o *e-value* que corresponde à probabilidade de se obter, com outra sequência aleatória de mesmo tamanho e composição de letras, outro alinhamento com *score* igual ou superior. O processo de alinhamento ocorre de forma aleatória e sequências sem alguma relação podem ser alinhadas. Sendo assim, quanto mais próximo de zero for o *e-value*, mais confiável será a consulta (ALTSCHUL *et al.*, 1990).

A bioinformática tem uma importância fundamental na agilização desse processo, permitindo a análise de múltiplos dados simultaneamente, usando algoritmos complexos e auxiliando no melhor entendimento de diversas doenças de base genética, principalmente as de heranças multifatorial e poligênica como ocorre com a DA. (PEREZ-IRATXETA *et al.*, 2002).

2. Justificativas

A realização de pesquisas sobre a fisiopatologia da DA em nosso meio é de importância estratégica, haja vista o aumento progressivo e contínuo da incidência desta devastadora doença em todo o mundo, inclusive no Brasil.

Pouco se conhece sobre os fatores de risco genéticos para a DA na população Brasileira. Além disso, os resultados deste estudo poderão ajudar a elucidar os mecanismos de manutenção da integridade dos sistemas biológicos relacionados com a capacidade de memorização em humanos.

Através do estudo das vias imune e inflamatórias na DA é provável que se possa identificar potenciais novos fatores genéticos de risco que auxilie no desenvolvimento de novas estratégias terapêuticas baseadas em modulações imune, tal como o uso de medicamentos antiinflamatórios e vacinas ao qual já sendo testadas (WISNIEWSKI *et al.*, 2008).

Posteriormente, estes dados serão relevantes para a prática do estudo, diagnóstico e planejamento de terapias para a DA na nossa população.

3. Objetivos

3.1. Gerais

Definir os principais genes e polimorfismos envolvidos na DA, priorizando genes ligados ao sistema imunológico, através de técnicas de Bioinformática.

3.2. Específicos

- Desenvolver uma nova metodologia que envolva ferramentas disponíveis na web, para o *screening* e validação de SNPs a partir de ESTs.
- Identificar novos potenciais polimorfismos genéticos de risco na Doença de Alzheimer em bancos de dados de *Microarrays*.

4. CAPÍTULO I

ARTIGO I

Linkage studies in familial idiopathic basal ganglia calcification: Separating the wheat from the chaff. **Am J Med Genet B Neuropsychiatr Genet.** 2008 Jul 28. In Press. DOI 10.1002/ajmg.b.30839.

Letter to the Editor

Linkage Studies in Familial Idiopathic Basal Ganglia Calcification: Separating the Wheat From the Chaff

J.R.M. Oliveira^{1,2*} and R.R. Lemos²¹Department of Neuropsychiatry, Federal University of Pernambuco (UFPE), Recife, PE, Brazil²Exzo Asami laboratory (LIKA), Federal University of Pernambuco (UFPE), Recife, PE, Brazil

Idiopathic Basal Ganglia Calcification (IBGC) is a neuropsychiatric condition characterized by brain calcinosis, heterogeneous motor impairment and behavioral symptoms. The IBGC1 locus was the first region linked to this phenotype in an American family, but another kindred from Spain was also reported as possibly associated with this locus. Our group excluded this locus in additional families together with an independent study of an Australian pedigree with IBGC, but without clinical symptoms. Recently, a large Italian family from a population isolate was excluded from IBGC1. However, there are unusual aspects concerning this Tyrolean family, especially if we consider that almost all the clinically affected subjects manifested symptoms and signs suggestive of a dysmorphic syndrome, associated with neuropsychiatric symptoms. Curiously, some of the clinical features in this kindred match with the autosomal dominant chromosomal instability syndrome reported in Japan. Previous studies show that the definition of an autosomal dominant pattern of inheritance is an assumption that might be considered cautiously in familial IBGC, due to the limited clinical penetrance for the brain calcifications and especially when there is no access to all the parents neuroimaging data. Families from an Italian isolate, such as Tyrol, with high inbreeding rates, are more likely to manifest recessive syndromes. Nevertheless, the current debate regarding the nosology of this heterogeneous phenotype demands the establishment of standard diagnostic criteria. The current identification of loci or mutations responsible for IBGC might help to elucidate this intriguing neuropsychiatric condition.

© 2008 Wiley-Liss, Inc.

KEY WORDS: IBGC; Fahr's disease; basal ganglia calcification; linkage; chromosome 14

Please cite this article as follows: Oliveira JRM, Lemos RR. 2008. Linkage Studies in Familial Idiopathic Basal Ganglia Calcification: Separating the Wheat From the Chaff. *Am J Med Genet Part B* 9999:1-2.

*Correspondence to: J.R.M. Oliveira, Av Prof Moraes Rego, Cidade Universitária, Recife, Pernambuco 50670-901, Brazil. E-mail: joao.ricardo@ufpe.br

Received 27 May 2008; Accepted 30 June 2008

DOI 10.1002/ajmg.b.30839

Published online 00 Month 2008 in Wiley InterScience (www.interscience.wiley.com)

© 2008 Wiley-Liss, Inc.

Dear Editor,

There has been extensive investigation into the genetic background of Familial idiopathic basal ganglia calcification (IBGC), a neuropsychiatric condition characterized by heterogeneous manifestations of motor impairment and behavioral disturbances that is inherited in an autosomal dominant manner and often reported as "Fahr's Disease" [Baba et al., 2005].

Recently, Volpato et al. [2008] reported a linkage study excluding a 14q chromosome region (locus IBGC1) in a large family with IBGC, originally from a population isolate in South Tyrol, Italy.

The IBGC1 locus was initially described for one large multigenerational American family in a 13.3 cM critical region (14q11.2-21.3). In addition to calcifications in the basal ganglia and extra-striatal regions, manifestations in this family included dysphagia, focal dystonia, tremor, parkinsonism, and psychosis [Geschwind et al., 1999].

Another small kindred, from Spain, was also reported as possibly linked to this locus, narrowing the candidate region to 10.9 cM. However, we excluded this locus in other families, together with an independent study of an Australian pedigree with IBGC but without neurological, cognitive and psychiatric symptoms [Brodaty et al., 2002; Oliveira et al., 2004].

However, it is intriguing to consider that almost all the clinically affected subjects from the Tyrolean family manifested symptoms and signs suggestive of a dysmorphic syndrome, associated with neuropsychiatric symptoms. One affected subject, with calcifications and dysmorphic features, presented no other symptoms and the majority of subjects with brain calcification were asymptomatic.

Previous neuroradiological analysis of brain calcinosis showed that some individuals might have no neurological symptoms despite the presence of extensive neuroimaging findings but studies measuring the total volume of calcification using an Electronic Planimeter and Coordinate Digitizer suggest a significantly greater amount of calcification in symptomatic individuals compared to asymptomatic subjects [Manyam et al., 2001; Manyam, 2005].

Curiously, some of the clinical features in this kindred match with the autosomal dominant chromosomal instability syndrome reported in Japan which includes mental retardation, hyperreflexia, ataxia, dysarthria, psychiatric findings, short stature, osteoporosis, and dental caries [Baba et al., 2005].

Volpato et al. [2008] should consider the possibility of facing a new syndrome that must be screened initially with a sensitive cytotype study, including microdeletion analysis and testing hypersensitivity to radiation and bleomycin. This could eventually explain the whole clinical heterogeneity in this pedigree and discard the possibility of comorbidity.

Additionally, a prior Japanese study showed that the definition of an autosomal dominant pattern of inheritance is an assumption that should be considered cautiously in familial IBGC when there is no access to all the parents neuroimaging

data, due to the limited clinical penetrance for the brain calcifications [Kobari et al., 1997]. Actually, families from an isolate, such as South Tyrol, where inbreeding rates are high, are more likely to manifest recessive syndromes.

The search for genetic markers related to IBGC is quite active and will be determinant for the definition of its pathophysiology. However, the current debate about the nosology of this heterogeneous phenotype demands a for the establishment of standard diagnostic criteria and the identification of loci or mutations responsible for IBGC might help to elucidate this intriguing neuropsychiatric condition [Schwarzbraun et al., 2004; Oliveira et al., 2007].

REFERENCES

- Baba Y, Broderick DF, Uitti RJ, Hutton ML, Wszolek ZK. 2005. Hereditary familial brain calcinosis syndrome. *Mayo Clin Proc* 80:641–651.
- Brodsky H, Mitchell P, Luscombe G, Kwok JJ, Badenhop RF, McKenzie R, Schofield PR. 2002. Familial idiopathic basal ganglia calcification (Fahr's disease) without neurological, cognitive and psychiatric symptoms is not linked to the IBGC1 locus on chromosome 14q. *Hum Genet* 110:8–14.
- Geschwind DH, Loginov M, Stern JM. 1999. Identification of a locus on chromosome 14q for idiopathic basal ganglia calcification (Fahr disease). *Am J Hum Genet* 65:764–772.
- Kobari M, Nogawa S, Sugimoto Y, Fukuuchi Y. 1997. Familial idiopathic brain calcification with autosomal dominant inheritance. *Neurology* 48:645–649.
- Manyam BV. 2006. What is and what is not "Fahr's disease". *Parkinsonism Relat Disor* 11:73–80.
- Manyam BV, Walters AS, Narla KR. 2001. Bilateral striopallidodentate calcinosis: Clinical characteristics of patients seen in a registry. *Mov Disord* 16:258–264.
- Oliveira JRM, Spiteri E, Sobrido MJ, Hopfer S, Klepper J, Voit T, Gilbert J, Wszolek ZK, Calne DB, Stoessl AJ, Hutton M, Manyam BV, Boiler F, Baquero M, Geschwind DH. 2004. Genetic heterogeneity in familial idiopathic basal ganglia calcification ("Fahr's disease"). *Neurology* 63:2165–2167.
- Oliveira JRM, Sobrido MJ, Spiteri E, Hopfer S, Meroni G, Petek E, Baquero M, Geschwind DH. 2007. Analysis of candidate genes at the IBGC1 locus associated with idiopathic basal ganglia calcification ("Fahr's Disease"). *J Mol Neurosci* 33:151–154.
- Schwarzbraun T, Vincent JB, Schumacher A, Geschwind DH, Oliveira J, Windpassinger C, Ofner L, Ledinegg MK, Kroisel PM, Wagner K, Petek E. 2004. Cloning, genomic structure, and expression profiles of TULIP1 (GARNL1), a brain-expressed candidate gene for 14q13-linked neurological phenotypes, and its murine homologue. *Genomics* 84:577–586.
- Volpato CB, De Grandi A, Buffone E, Pichler I, Gebert U, Schifferle G, Schönhuber R, Pramstaller PP. 2008. Exclusion of *Linkage*^{q2} to chromosome 14q in a large south tyrolean family with idiopathic basal ganglia calcification (IBGC). *Am J Med Genet Part B* 10.1002/ajmg.b.30748.

Q1: Please check the affiliations.

Q2: Please provide the volume number and page range.

5. CAPÍTULO II

ARTIGO II

The Genetics of Alzheimer's Disease in Brazil: 10 Years of Analysis in a Unique Population. **J Mol Neurosci.** v. 37, 74-79, 2009.

The Genetics of Alzheimer's Disease in Brazil: 10 Years of Analysis in a Unique Population

J. R. M. Oliveira · A. L. Nishimura ·
R. R. Lemos · M. Zatz

Received: 25 March 2008 / Accepted: 11 June 2008 / Published online: 8 July 2008
© Humana Press 2008

Abstract Alzheimer's Disease (AD) is the most common type of dementia among the elderly, with devastating consequences for the patient, their relatives, and caregivers. More than 300 genetic polymorphisms have been involved with AD, demonstrating that this condition is polygenic and with a complex pattern of inheritance. This paper aims to report and compare the results of AD genetics studies in case-control and familial analysis performed in Brazil since our first publication, 10 years ago. They include the following genes/markers: Apolipoprotein E (APOE), 5-hydroxytryptamine transporter length polymorphic region (5-HTTLPR), brain-derived neurotrophin factor (BDNF), monoamine oxidase A (MAO-A), and two simple-sequence tandem repeat polymorphisms (DXS1047 and D10S1423). Previously unpublished data of the interleukin-1 α (IL-1 α) and interleukin-1 β (IL-1 β) genes are reported here briefly. Results from others Brazilian studies with AD patients are also reported at this short review. Four local families studied with various markers at the chromosome 21, 19, 14, and 1 are briefly reported for the first time. The importance of studying DNA samples from Brazil is highlighted because of the uniqueness of its population, which presents

both intense ethnical miscegenation, mainly at the east coast, but also clusters with high inbreeding rates in rural areas at the countryside. We discuss the current stage of extending these studies using high-throughput methods of large-scale genotyping, such as single nucleotide polymorphism microarrays, associated with bioinformatics tools that allow the analysis of such extensive number of genetics variables, with different levels of penetrance. There is still a long way between the huge amount of data gathered so far and the actual application toward the full understanding of AD, but the final goal is to develop precise tools for diagnosis and prognosis, creating new strategies for better treatments based on genetic profile.

Keywords Alzheimer's disease · Polymorphisms · Genetic risk factors · Brazil

Introduction

Recently, the plethora of data produced at the field of genetic risk factors for Alzheimer's Disease (AD) was compiled and summarized in a new public database entitled ALZGENE. This website centralizes and curates most of the genotyping data generated so far from AD case-control analysis and might be used especially for meta-analysis studies (<http://www.alzforum.org>).

This was a key initiative, considering the amount of articles published about this issue, especially during the last 15 years.

More than 300 genetic polymorphisms have been involved with AD, but only a few were independently confirmed, demonstrating that this condition is polygenic and with a complex pattern of genetic inheritance (Bertram et al. 2007).

J. R. M. Oliveira (✉)
Department of Neuropsychiatry,
Federal University of Pernambuco,
Recife, PE, Brazil
e-mail: joao.ricardo@ufpe.br

J. R. M. Oliveira · R. R. Lemos
Keizo Asami Laboratory (LIKA),
Federal University of Pernambuco,
Recife, PE, Brazil

A. L. Nishimura · M. Zatz
The Human Genome Study Center, University of São Paulo,
São Paulo, SP, Brazil

When our group started studying this issue, the foundations of molecular studies in AD were based on specific studies of families with early-onset Alzheimer's disease (EOAD) and late-onset Alzheimer's disease (LOAD). Mutations at the amyloid precursor protein (APP) and at the presenilin genes (PS1 and PS2) were already reported and the apolipoprotein E $\epsilon 4$ allele (ApoE $\epsilon 4$) was considered the most important polymorphism involved mainly in LOAD (Eastwood et al. 1996).

Around a decade ago, the first case-control studies and family analysis were performed in Brazilian patients with some genes previously linked to AD, such as APOE, 5HTTLPR, DXS1047, D10S1423, BDNF, and MAO-A (Oliveira et al. 1997, 1998, 1999a, b; Nishimura et al. 2000, 2001, 2004, 2005). Previously unpublished data of the interleukin-1 α (IL-1 α) and interleukin-1 β (IL-1 β) genes are also reported here together with a brief report of markers at the chromosomes 21, 19, 14, and 1 in a few families with AD.

The Brazilian population has various interesting genetic idiosyncrasies, initially characterized 50 years ago. The groups living at the coast are highly mixed, with low inbreeding rates and presenting a major genetic contribution from Native Indians, Europeans, and African Americans. On the other hand, the population living in rural areas presents some of the highest inbreeding rates reported so far (Freire-Maia 1957).

We focused our recruitment at the coast population in two major cities: Recife and São Paulo. Performing case-control studies in these groups with low inbreeding rates is crucial due to the possibility of finding risk factors that are widely relevant in various ethnicities and not only limited to a specific population because the overrepresentation of rare variants in specific ethnic groups may confuse interpretation of association analyses (Keen-Kim et al. 2006).

After the publication from various other research groups studying the same mutations and also other polymorphisms, it is clear that the genetic background for AD is polygenic and complex.

A more detailed analysis of our findings will follow along this article.

Case-control Studies

In general, the genetic background of the population recruited in our case-control studies presented a distribution of 3:1 for Caucasoids and African/Brazilians, respectively. This disproportioned distribution decreased the number of subject available to correlation studies between the ethnical backgrounds and specific genetic entity.

The sample sizes were progressively increased since 1997 and the estimated allele frequencies were consistently checked for the Hardy-Weinberg equilibrium.

Studying AD patients and another group with memory deficit, but not demented (pre-clinical subjects), we found significant differences at the APOE $\epsilon 4$ allele distribution between AD patients and controls. This study also verified a significant association between the $\epsilon 4$ allele and the pre-clinical subjects with positive family history for dementia, confirming this polymorphism as a susceptibility factor and suggesting that it might be useful for screening risk groups (Oliveira et al. 1997). This result coincides with the majority of studies that bestowed to the APOE $\epsilon 4$ allele the status of the most important risk factor for AD (Bertram et al. 2007).

The polymorphism at the promoter region of the serotonin transporter gene (5-HTTLRP), an insertion/deletion of 44 base pairs generating two alleles with the short version (s) decreasing the gene transcriptional activity was widely studied in various neuropsychiatric conditions and personality traits (Canli and Lesch 2007). The short version was also considered a risk factor to develop AD in a Brazilian group of AD patients (Oliveira et al. 1998). Curiously, we did not find significant association for risk increase when studying Brazilian patients with both s variant and $\epsilon 4$ allele (Oliveira et al. 1999a).

This study confirmed for the first time the findings of Li et al. (1997), but later other studies in Japanese population found no statistically significant association between AD and the short allele (Kunugi et al. 2000).

Nevertheless, studies with polymorphisms related with serotonin, involving receptors, transporters, and enzymatic machinery, enhanced our knowledge about specific pathological mechanisms involved at this neuropsychiatric condition (Oliveira and Zatz 1999b).

Simultaneously, an important Genome wide scan found that the 202 bp allele in locus DXS1047 and 234 bp in locus D10S1423 were associated with AD. These researchers had observed that the patients with these two alleles presented reduced levels of dopamine in six cortical regions and high concentrations of cortical norepinephrine, suggesting a dosage effect of the alleles in question (Zubenko et al. 1998).

Both locus DXS1047 and D10S1423 were analyzed in Brazilian patients, but no statistically significant association was found (Nishimura et al. 2000, 2001).

The C-270T polymorphism, located at the 5' non-coding region of brain-derived neurotrophin gene (BDNF) was also associated with AD and the patients with the T variant present an increase of the susceptibility to develop the illness. In accordance with these authors, the polymorphism can modify the expression of protein BDNF, being capable of relating to the pathogenesis of AD (Kunugi et al. 2001).

However, we found no significant association between these polymorphisms and the AD (Nishimura et al. 2004).

Monoamine oxidase A (MAOA) is a mitochondrial enzyme that catalyzes the degradation of different amines including neurotransmitters like dopamine, norepinephrine, and serotonin. Several studies have shown that MAOA plays important role in human behavior and physiology, being associated with mild mental retardation and impulsive aggressive behavior, mood disorders, and alcoholism with antisocial personality (Sabol et al. 1998).

We found no association between the locus of MAOA and LOAD, but curiously, a combination of the MAOA allele 1, the 5HTTLPR short variant and APOE $\epsilon 4$ was more frequent in patients, reinforcing the hypothesis that different genes might act together to modulate the disease (Nishimura et al. 2005).

Other studies have attempted to find a relationship between inflammatory pathways and psychiatric disorders. Polymorphisms of the interleukin-1 (IL-1) gene complex such as IL-1 α and IL-1 β have been associated to different neuropsychiatric conditions, including AD (Katila et al. 1999; Du et al. 2000; Mattila et al. 2002; Fertuzinhos et al. 2004).

To test the association of these polymorphisms and AD, a total of 195 LOAD Brazilian patients with mean age of 68.7 ± 8 years were selected in accordance with the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria. The 188 age-matched controls with mean age of 72.3 ± 9.75 were selected according to the Mine Mental State Examination (MMSE) and/or Blessed Scale, socio-cultural and comparable ethnic background. All patients and healthy subjects gave informed consent for participation in this study with the approval of the ethical committee of the Institute of Biosciences, University of São Paulo.

In order to evaluate the distribution of IL-1 polymorphisms in the Brazilian population, genomic DNA was isolated from peripheral blood according to standard procedures and the genotyping was based on previous study (Katila et al. 1999).

No significant association was observed when we compared either allele or genotype distribution for IL-1 α and IL-1 β polymorphisms ($P=0.813$ and $P=0.607$) for allelic and for genotype distribution ($P=0.825$ and $P=0.427$), respectively, suggesting they might not confer an increased risk for LOAD in the Brazilian population.

A few other Brazilian groups also performed similar studies, examining the association of the methylenetetrahydrofolate reductase enzyme (MTHFR) with AD. MTHFR is a central enzyme of the folate metabolism forming the substrate needed for the transferring reaction, an important co-factor to homocysteine (Hcy) methylation. Epidemiological studies have reported an association between increased plasma Hcy level and AD (Fernandez and Scheibe 2005; Religa et al. 2003).

Da Silva et al. (2006) found no association between MTHFR polymorphisms and AD patients from Brazil, but found a difference at the level of expression at the blood serum.

Recently, meta-analyses aimed for this polymorphism did not confirm it as a major risk factor for the AD (Bertram et al. 2007).

All these Brazilian case-control studies are summarized in Table 1.

Family Studies

The study of families with LOAD is extremely difficult due to the fact that potential affected subjects will decrease from

Table 1 A summarized analysis of the Brazilian case-control studies

References	Polymorphisms	P value	Sample size
Oliveira et al. (1997)	ApoE- $\epsilon 4$	$P < 0.05$	57
Oliveira et al. (1998)	5-HTTLPR	$P < 0.05$	81
Nishimura et al. (2000)	DXS1047	$P > 0.05$	130
Nishimura et al. (2001)	D10S1423	$P > 0.05$	130
Nishimura et al. (2004)	BDNF (C-270T)	$P > 0.05$	188
	BDNF (C-270T) + ApoE- $\epsilon 4$	$P < 0.05$	
Nishimura et al. (2005)	MAOA	$P > 0.05$	128
	MAOA + ApoE- $\epsilon 4$ + 5-HTTLPR	$P < 0.05$	
Unpublished data	IL-1 α and IL-1 β	$P > 0.05$	195
	ApoE- $\epsilon 4$	$P < 0.05$	
Fernandez and Scheibe (2005)	MTHFR (C-677T)	$P > 0.05$	30
	ApoE- $\epsilon 4$	$P < 0.05$	
da Silva et al. (2006)	MTHFR (C-677T)/(A-1298C)	$P > 0.05$	43

various age-related causes before developing the dementia symptoms, so the number of affected subjects available for Linkage analysis and association studies is often scarce. This is why the most elucidative family studies came from the analysis of EOAD (Shen and Raymond 2007).

We were able to recruit four small kindred with LOAD and tested the pattern of segregation with markers for the chromosomes already involved with familial AD, such as D1S158, D1S479, F13B, D1S53, D1S179, D1S217, D1S249, D1S175, D1S178, D14S52, D14S53, D14S42, N1411, D21S120, D21S13E, APP, APOCII (MFD5), D19S278, D19S178, D19S224, D19S47, and D19S75. None of these markers showed a distinct and significant segregation pattern across the families based on Linkage analysis. The APOE and 5-HTTLPR were also genotyped at these patients to check possible association. Curiously, these were the only markers associated with the affection status (unpublished data).

Other Genetic Studies at the Brazilian Population

Cytogenetic studies were also performed at the Brazilian population in order to find the major differences between AD patients and controls. Kormann-Bortolotto et al. (1992) examined chromosomal fragility in both AD and elderly control groups and found a higher frequency of fragile sites in 6p21, but not in the young controls. These caryotype bands overlap with the position of the MTBT2 and MTBT1 genes, which are microtubule (beta)-associated protein tau-like and tau 1, respectively.

The maturation rates of the 28S/18S subunits of rRNA were analyzed in order to establish an association of this event with the AD. Payão et al. (1998) and da Silva et al. (2000) found reduced maturation and transcription rate, suggesting a possible change of these processes or preference degradation of the 28S subunit.

By analyzing the maturation rate the subunits 28S/18S and the effect of the allele APOE e4, no association was found with AD, suggesting that the two different events are involved in the etiology of the AD, but traveling through different paths (Tavares et al. 2004).

Another genetic study checked for association at the Werner helicase polymorphism and AD, but the result was negative (Payão et al. 2004).

Possible Reasons to Explain the Results Disparity in Brazilian Studies

Over the 10 years of research with the Brazilian population, eight polymorphic markers were addressed by our group; however, only two showed statistically significant differences of distribution in case-control studies.

Some possible reason to explain our results when comparing with positive studies are the relative reduced sample size and clinical diagnostic criteria.

Our patients were diagnosed with probable/possible AD, and we did not have access to brain tissue. This procedure increases the chances of including patients with other types of dementia. However, the positive findings with the APOE analysis confirm that we were working with a population well selected enough to detect the significance similar to most of the studies performed worldwide.

Another possible explanation for this result would be the great ethnical variability in the Brazilian population that could mask the relevance of some markers in a few specific groups or that other susceptibility genes would be acting as risk factors at the Brazilian population (NCI-NHGRI Working Group on Replication in Association Studies 2007).

For various reasons, case-control studies testing allelic association might produce positive and negative results in different populations also due to different ethnical background, and this is a good reason to compare data from different groups of patients (Keen-Kim et al. 2006).

Controversial results are found when different analyses are performed in complex disorders from various populations worldwide (Freimer and Sabatti 2007).

New Perspectives Using High-throughput Methods of Genotyping Genetic Variations, Such as Single-nucleotide Polymorphism (SNPs) Microarrays

The availability of straightforward techniques to perform high-throughput genotyping, by using SNPs microarrays, multiplex and real-time polymerase chain reaction (PCR) will provide faster and more precise methods to extend the studies genetic risk factor for AD at this same population (Pearson et al. 2007).

Given the large number of candidate genes that have already been identified for AD, (<http://www.alzforum.org>), the most recent studies focus on understanding how these genes interact with each other to trigger the disease. Analysis of microarray suits this purpose, pointing differences in the levels of gene expression on a large scale or, more recently, identifying variations SNPs in LOAD. The different approaches with the same technique specify groups of genes involved in the pathological process (Colangelo et al. 2002; Hamshere et al. 2007).

The bioinformatics has a fundamental importance in speeding up that process, allowing the analysis of multiple data simultaneously, using complex algorithms to help the better understanding of various complex disorders, such as AD (Perez-Iratxeta et al. 2002).

There is still a long way between the huge amount of data gathered so far and the actual application toward the

full understanding of AD, but the final goal is to develop precise tools for diagnosis and prognosis, creating new strategies for better treatments based on genetic profile.

References

- Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., & Tanzi, R. E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature Genetics*, 39, 17–23. doi:10.1038/ng1934.
- Canli, T., & Lesch, K. P. (2007). Long story short: the serotonin transporter in emotion regulation and social cognition. *Nature Neuroscience*, 10, 1103–1109. doi:10.1038/nn1964.
- Colangelo, V., Schurr, J., Ball, M. J., Pelaez, R. P., Bazan, N. G., & Lukiw, W. J. (2002). Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *Journal of Neuroscience Research*, 70, 462–473. doi:10.1002/jnr.10351.
- da Silva, A. M., Payão, S. L., Borsatto, B., Bertolucci, P. H., & Smith, M. A. (2000). Quantitative evaluation of the rRNA in Alzheimer's disease. *Mechanisms of Ageing and Development*, 120, 57–64. doi:10.1016/S0047-6374(00)00180-9.
- da Silva, V. C., Ramos, F. J., Freitas, E. M., et al. (2006). Alzheimer's disease in Brazilian elderly has a relation with homocysteine but not with MTHFR polymorphisms. *Arquivos de Neuro-Psiquiatria*, 64, 941–945.
- Du, Y., Dodel, R. C., Eastwood, B. J., et al. (2000). Association of an interleukin 1 alpha polymorphism with Alzheimer's disease. *Neurology*, 55, 480–483.
- Eastwood, R., Amaducci, L., Brayne, C., et al. (1996). The challenge of the dementias. *Lancet*, 374, 1303–1307.
- Fernandez, L. L., & Scheibe, R. M. (2005). Is MTHFR polymorphism a risk factor for Alzheimer's disease like APOE? *Arquivos de Neuro-Psiquiatria*, 63, 1–6. doi:10.1590/S0004-282X2005000100001.
- Fertuzinhos, S. M., Oliveira, J. R., Nishimura, A. L., et al. (2004). Analysis of IL-1alpha, IL-1beta, and IL-1RA [correction of IL-RA] polymorphisms in dysthymia. *Journal of Molecular Neuroscience*, 22, 251–256. doi:10.1385/JMN:22:3:251.
- Freimer, N. B., & Sabatti, C. (2007). Human genetics: variants in common diseases. *Nature*, 22, 828–830. doi:10.1038/nature05568.
- Freire-Maia, N. (1957). Inbreeding in Brazil. *American Journal of Human Genetics*, 9, 284–298.
- Hamshere, M. L., Holmans, P. A., Avramopoulos, D., et al. (2007). Genome-wide linkage analysis of 723 affected relative pairs with late-onset Alzheimer's disease. *Human Molecular Genetics*, 16, 2703–2712. doi:10.1093/hmg/ddm224.
- Katila, H., Hänninen, K., & Hurme, M. (1999). Polymorphisms of the interleukin-1 gene complex in schizophrenia. *Molecular Psychiatry*, 4, 109–111. doi:10.1038/sj.mp.4000483.
- Keen-Kim, D., Mathews, C. A., Reus, V. L., et al. (2006). Overrepresentation of rare variants in a specific ethnic group may confuse interpretation of association analyses. *Human Molecular Genetics*, 15, 3324–3328. doi:10.1093/hmg/ddi408.
- Kormann-Bortolotto, M. H., de Smith, M. A., & Toniolo Neto, J. (1992). Fragile sites, Alzheimer's disease, and aging. *Mechanisms of Ageing and Development*, 65, 9–15. doi:10.1016/0047-6374(92)90122-T.
- Kunugi, H., Ueki, A., Otsuka, M., et al. (2000). Alzheimer's disease and 5-HTTLPR polymorphism of the serotonin transporter gene: no evidence for an association. *American Journal of Medical Genetics*, 96, 307–309. doi:10.1002/1096-8628(20000612)96:3<307::AID-AJMG14>3.0.CO;2-O.
- Kunugi, H., Ueki, A., Otsuka, M., et al. (2001). A novel polymorphism of the brain-derived neurotrophic factor (BDNF) gene associated with late-onset Alzheimer's disease. *Molecular Psychiatry*, 6, 83–86. doi:10.1038/sj.mp.4000792.
- Li, T., Holmes, C., Sham, P. C., et al. (1997). Allelic functional variation of serotonin transporter expression is a susceptibility factor for late onset Alzheimer's disease. *Neuroreport*, 10, 683–686.
- Mattila, K. M., Rinne, J. O., Lehtimäki, T., Roytta, M., Ahonen, J. P., & Hurme, M. (2002). Association of an interleukin 1B gene polymorphism (–511) with Parkinson's disease in Finnish patients. *Journal of Medical Genetics*, 39, 400–402. doi:10.1136/jmg.39.6.400.
- N.C.I.-N.H.G.R.I., Working Group on Replication in Association Studies (2007). Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS. Replicating genotype-phenotype associations. *Nature*, 447(7145), 655–660. doi:10.1038/447655a.
- Nishimura, A. L., Guindalini, C., Oliveira, J. R. M., et al. (2005). Monoamine oxidase A polymorphism in Brazilian patients: risk factor for late onset Alzheimer's disease? *Journal of Molecular Neuroscience*, 7, 213–217. doi:10.1385/JMN:7:2:213.
- Nishimura, A. L., Oliveira, J. R. M., Maitoli, S., et al. (2000). Analysis of the disease risk loci DXS1047 polymorphism in Alzheimer Brazilian patients. *Molecular Psychiatry*, 5, 563–566. doi:10.1038/sj.mp.4000767.
- Nishimura, A. L., Oliveira, J. R. M., Mitne-Neto, M., et al. (2004). Lack of association between the brain-derived neurotrophin factor (C-270T) polymorphism and Late-Onset Alzheimer's Disease (LOAD) in Brazilian patients. *Journal of Molecular Neuroscience*, 22, 257–260. doi:10.1385/JMN:22:3:257.
- Nishimura, A. L., Oliveira, J. R. M., Otto, P. A., et al. (2001). No evidence of association between the D10S1423 locus and Alzheimer disease in Brazilian patients. *Journal of Neural Transmission*, 108, 305–310. doi:10.1007/s007020170076.
- Oliveira, J. R. M., Gallindo, R. M., Maia, L. G. S., et al. (1998). The short variant of the polymorphism within the promoter region of the serotonin transporter gene is a risk factor for late onset Alzheimer's disease. *Molecular Psychiatry*, 3, 438–441. doi:10.1038/sj.mp.4000417.
- Oliveira, J. R. M., Lima-Filho, J. L., Shimokomaki, C. M., et al. (1997). The use of apolipoprotein E genotype for preclinical detection of risk's group for Alzheimer's disease. *American Journal of Medical Genetics Neuropsychiatric Genetics*, 74, 216–217. doi:10.1002/(SICI)1096-8628(19970418)74:2<216::AID-AJMG22>3.0.CO;2-K.
- Oliveira, J. R. M., Shimokomaki, C. M., Brito-Marques, P. R., et al. (1999a). The association of the short polymorphism of the 5-HTTLPR polymorphism and the ApoE ε4 allele does not increase the risk for late onset Alzheimer's Disease. *Molecular Psychiatry*, 4(1), 19–20.
- Oliveira, J. R. M., & Zatz, M. (1999b). The study of genetic polymorphisms related to serotonin in Alzheimer's disease: a new perspective in a heterogenic disorder. *Brazilian Journal of Medical and Biological Research*, 32, 463–467.
- Payão, S. L., de Labio, R. W., Gatti, L. L., Rigolin, V. O., Bertolucci, P. H., & de Smith, M. A. (2004). Werner helicase polymorphism is not associated with Alzheimer's disease. *Journal of Alzheimer's Disease*, 6, 591–594.
- Payão, S. L., Smith, M. A., Winter, L. M., & Bertolucci, P. H. (1998). Ribosomal RNA in Alzheimer's disease and aging. *Mechanisms*

- of Ageing and Development*, 105, 265–272. doi:10.1016/S0047-6374(98)00095-5.
- Pearson, J. V., Huentelman, M. J., Halperin, R. F., et al. (2007). Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. *American Journal of Human Genetics*, 80, 126–139. doi:10.1086/510686.
- Perez-Iratxeta, C., Bork, P., & Andrade, M. (2002). Association of genes to genetically inherited diseases using data mining. *Nature Genetics*, 31, 316–319.
- Religa, D., Stycz, M., Peplonska, B., et al. (2003). Homocysteine, apolipoprotein E and methylenetetrahydrofolate reductase in Alzheimer's disease and mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, 16, 64–70. doi:10.1159/000070677.
- Saboi, S. Z., Hu, S., & Hamer, D. (1998). A functional polymorphism in the monoamine oxidase A gene promoter. *Human Genetics*, 103, 273–279. doi:10.1007/s004390050816.
- Shen, J., & Raymond, J. K. (2007). The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 403–409. doi:10.1073/pnas.0608332104.
- Tavares, W. M., Sperança, M. A., de Lábio, R. W., et al. (2004). Apolipoprotein E4 allele and ribosomal genes in Alzheimer's disease. *Journal of Alzheimer's Disease*, 6, 391–395.
- Zubenko, G. S., Hughes, H. B., Stiffler, J. S., Hurtt, M. R., & Kaplan, B. B. (1998). A genome survey for novel Alzheimer disease risk loci: results at 10-cM resolution. *Genomics*, 50, 121–128. doi:10.1006/geno.1998.5306.

6. CAPÍTULO III

ARTIGO III

In silico identification of new genetic variations as potential risk factors for Alzheimer's Disease in a Microarray oriented simulation. **J Mol Neurosci.** 2009, In Press: DOI 10.1007/s12031-009-9191-x.

In Silico Identification of New Genetic Variations as Potential Risk Factors for Alzheimer's Disease in a Microarray-oriented Simulation

R. R. Lemos · CH Castelletti · J. L. Lima Filho ·
E. T. Marques · J. R. M. Oliveira

Received: 19 November 2008 / Accepted: 24 February 2009
© Humana Press 2009

Abstract Genomic and proteomic studies of neurodegenerative disorders require complementary approaches to integrate the massive amount of data generated in high throughput experimental procedures. We propose a Bioinformatics pipeline in which expression studies guide the selection of candidate genes that should be screened for

potential new genetic variations from a public expressed site tags (ESTs) database. Motivated by the former interest of our group in genetic polymorphisms involved with the immune system, we selected five genes from a previous expression microarrays study of hippocampal cornu ammonis (CA1) area of Alzheimer's Disease subjects (AD). The CLCbio Workbench Combined® version 3.6.2. was initially used to build ESTs and mRNA files retrieved respectively from the Goldenpath of University of California Santa Cruz (UCSC) and National Center for Biotechnology Information (NCBI) databases and latter to perform multiple batches of Smith-Waterman alignments. A total of 116 ESTs sequences were selected after proper stringent parameters were applied to the first set of mismatches. The annotation revealed various classes of variations, most of them deletions (194), but also transitions (three), transversions (52), synonymous (46), non-synonymous (44), and Single Nucleotide Polymorphisms (SNPs) in Untranslated Regions (UTRs) (36). Deletions are often associated to major genetics syndromes with dysmorphic features. However, various recent studies show that common microdeletions might be highly associated with common neuropsychiatric disorders such as schizophrenia, autism, mental retardation, or even in various ethnicities, detected in whole genome sequencing experiments. A virtual validation confirmed that some of the variations identified were previously reported and confirmed in DNA samples, showing that this method is a feasible way to detect genetic variations that merit further exploration in AD genetic risk factor association studies.

R. R. Lemos · C. Castelletti · J. L. Lima Filho · J. R. M. Oliveira
Keizo Asami Laboratory (LIKA),
Federal University of Pernambuco,
Recife, PE, Brazil

E. T. Marques
Virology and Experimental Therapy Laboratory,
Aggeu Magalhães Research Center,
CPqAM/FIOCRUZ,
Recife, Brazil

E. T. Marques
Department of Medicine, Division of Infectious Diseases,
The Johns Hopkins School of Medicine,
Baltimore, MD, USA

E. T. Marques
Department of Pharmacology and Molecular Sciences,
The Johns Hopkins School of Medicine,
Baltimore, MD, USA

J. R. M. Oliveira
Department of Neuropsychiatry,
Federal University of Pernambuco,
Recife, PE, Brazil

J. R. M. Oliveira (✉)
Department of Neuropsychiatry,
Federal University of Pernambuco (UFPE),
50670901 Recife, PE, Brazil
e-mail: joao.ricardo@ufpe.br

Keywords In silico · Alzheimer's disease · Bioinformatics · ESTs · Microarrays

Introduction

Genomics and proteomics researchers who study the genetic risk factors for Alzheimer's disease (AD) need urgently an approach to integrate the massive amount of data that is generated in high throughput experimental procedures (Miller et al. 2008).

Large-scale experiments, such as expression and genotyping microarrays, may facilitate the identification and analysis of thousands of disease related nucleotide variations, including single nucleotide polymorphisms (SNPs), deletions and insertions of a few base pairs, microdeletions or even major chromosomal changes (Yu et al. 2002; Stefansson et al. 2008; Walsh et al. 2008).

Recently, Coppola et al. (2008) identified variations at the Progranulin gene in DNA samples of two outliers which presented noticeable down-regulation at this gene in comparison to previous expression microarrays performed during the same study.

The amount of genomic data available for data mining analysis in public and private databases is rapidly increasing. Expressed site tags (ESTs) are among the several types of sequences that are available and already largely explored, for sequencing purposes and for the isolation of specific subsets of mRNA in expression studies (Picoult-Newberg et al. 1999).

Bioinformatics tools were built with different purposes and are currently curated and improved at the same time that others are created to process ESTs, allowing the simultaneous overlapping and comparison of numerous batches of sequences in order to identify genetic variations such as single SNPs (Useche et al. 2001; Barker et al. 2003; Forment et al. 2008; Navratil et al. 2008; Tang et al. 2006, 2008).

Aiming to address this issue, we propose a Bioinformatics pipeline where previous AD expression studies, using microarrays, were used as a unique source of candidate genes that were scrutinized in public data bases in search of non previously reported variations that might

be explored afterwards to identify new genetic variations as potential risk factors for disease susceptibility.

Motivated by the previous interest of our group in genetic polymorphisms of the Interleukin-1 panel in AD and Mood Disorders, we selected genes from an expression microarray study in hippocampus of affected patients which infers that the up-regulation of IL-1 α , IL-1 β , COX-2, PLA2G4A, and CD83 genes could be due to the inflammatory process, neurotoxicity and cell death (Colangelo et al. 2002; Fortuzinhos et al. 2004; Oliveira et al. 2009).

Abundant evidence suggests that an inflammatory process follows the β -amyloid deposits and neurofibrillary tangles aggregates in the brain of AD patients, contributing to the degenerative process and disease progression (Heneka and O'Banion 2007).

The study of the immune and inflammatory pathway in AD is likely to identify potential new risk factors and to support the development of new therapeutic strategies based upon immune modulation, such as the use of anti-inflammatory medications and vaccines, which have already been tested (Wisniewski and Konietzko 2008).

Methodology

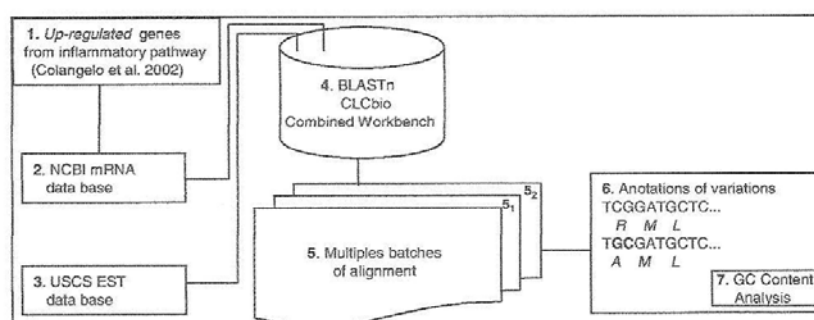
Raw Data Acquisition, Alignment and Annotation

An overview of the proposed pipeline is available in Fig. 1.

For this pilot analysis, five up-regulated genes were selected from a previous microarray study of hippocampal cornu ammonis 1 area in brain samples of AD patients (Colangelo et al. 2002): Interleukin 1 alpha (IL-1 α), Interleukin 1 beta (IL-1 β), Prostaglandin-endoperoxidase synthetase 2 (PTGS2 or COX-2), Phospholipase A2 group IV A (PLA2G4A) and Cluster of Differentiation 83 (CD83).

The software CLCbio Workbench Combined[®] version 3.6.2. was used during all the following steps. Initially, to build spliced ESTs and mRNA files by retrieving sequences

Figure 1 Sequential steps of the proposed Bioinformatics pipeline for screening candidate variations from ESTs databases



progression of AD neuro-pathological hallmarks (Minghetti 2004; Akiyama et al. 2000; Colangelo et al. 2002; Grimaldi et al. 2000; Nicoll et al. 2000; Tuppo and Arias 2005; Breloer and Fleischer 2008; Stephenson et al. 1996).

Additional expression studies showed that genes involved with immune responses are well-linked with AD and are widely up-regulated, indicating that inflammatory pathways is candidate for progression of disease (Blalock et al. 2004; Colangelo et al. 2002).

However, results of expression analysis showed serum IL-1 levels down-regulated in AD and pointed this protein and other 17 additional molecules as an important panel to predict which patients with mild cognitive impairment would progress to AD (Ray et al. 2007). It suggests that neuroimmune dysfunction present in AD might be limited to the CNS tissue, as evidenced by studies with hippocampal tissue (Blalock et al. 2004; Colangelo et al. 2002).

The annotation originated from our analysis revealed various classes of variations, most of them deletions (194), but also transitions (three), transversions (52), synonymous (46), non-synonymous SNPs (44) and SNPs in UTRs (36).

A search at the NCBI SNPs database identified that four variations detected were previously deposited and confirmed in DNA samples, suggesting that this method is a feasible approach to detect genuine genetic variations (see Fig. 2).

The genes analyzed in this study have been widely involved with the orchestrated modulation of immunological response, synaptic activity and memory consolidation. More recently, the IL-1 pathway has been particularly involved with the

a. CD83
205aa

1 2 3 4 5

BP296648 c.268delT Asp30fs
BP296531 c.268delT Asp30fs

BG70557 c.681delC Leu 168fs
B1758803 c.680delC Leu 168fs

AL575500 926_933delCATCTGGA UTR
BU162557 926_930delCATCT UTR

BEX30187 c.[628delT:649_650delAC] Leu151fs
BEX30134 c.[628delT:649_650delAC] Leu151fs

BE439432
BE619081 } 190 Stop Codon

BG121477 c.573delT Phe162fs
BE439432 c.573delT Phe162fs

185 Stop Codon

BG121477
ES309163
B1819099

b. IL-1β
269aa

1 2 3 4 5 6 7

B1907808 c.657delA 191delfs
BG250625 c.657delA 191delfs

220 Stop Codon

AI609005 BM999417
BG252492 CA307234

AI609005 c.700_701delAA 205delAsn+206Tyr>Leu
B1519707 c.700_701delAA 205delAsn+206Tyr>Leu

c. COX-2
604aa

1 2 3 4 5 6 7 8 9 10

DB019295 DA882455 DA277667
DN995052 BX363285 CN337323 } **I132G>G+I33G>C** UTR

DB223983 BP335898 BE717711 ALS83668 DB480063 DN995052 DB019295 BE785396
DA676889 DA170263 DA182079 DA88245 DA677048 DA67080 DA670627 DA486815
DA549596 DA499416 DA277667 DA674771 DA673470 DA676702 DA673400 DA675888
DA671756 DA674666 DA678032 DA675968 DA672971 DA672566 CV0023684 CN337323 } **c.440C>G**
CD609928 CB960307 BX363285 BP337788 BP335244 BP337860 BP336843 } Synonymous Mutation

d. PLA2G4A
749aa

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

c.2157A>G { CD104330; CB985872; T27656

UTR
CDS

Interestingly, in four genes, some mismatches were detected simultaneously when aligning sequences originally from different ESTs libraries, clustering some variations in specific regions. The CD83 gene presented this overlapping at the exons 2, 4 and 5. At the IL-1 β the mismatches clustered in Stop Codon regions and a single cluster was observed in PLA2G4A gene, exon 16. The largest clusters was detected at the COX-2 gene, exon 3.

This finding makes them actual promising candidates to be explored afterwards as candidate risk factors, since they were not derived from a single isolated EST sequence (see Fig. 2).

Besides of a wide set of sequences analysis tools, the UCSC genome browser platform displays and quantifies the "GC content" of the candidates regions and this is particularly important to identify regions more likely to generate sequencing artifacts. For example, a preliminary analysis showed low "GC content" in regions where deletions were predicted for the COX-2 gene.

Deletions ranging from 1 to 10 bp were the most common finding and present in coding regions and 5' or 3' UTRs region (Supplementary data). Curiously, Kidd et al. (2008) compared the whole genome of eight subjects and found 1.695 polymorphisms sites including: 747 deletions, 724 insertions, and 224 inversions.

Deletions are structural variations often associated with major genetic syndromes and dysmorphic features, howev-

er, recent studies of neuropsychiatric diseases such as schizophrenia, autism and mental retardation, found associations with microdeletions (Yu et al. 2002; Stefansson et al. 2008; Walsh et al. 2008).

We found a similar number of synonymous and non-synonymous mutations detected after translation analysis. Some non-synonymous mutations caused by microdeletions were predicted to change the coding frame and in some cases generating Stop Codon and truncated protein products (Table 1).

Current reports emphasize the relevance of frameshift caused by single deletions in multifactorial disorders and an interesting frameshift caused by mutation c.795delC in OGG1 gene was recently detected in AD patients, changing the coding frame and adding 59 amino acid residues downstream after terminator codon, causing reduced enzymatic activity (Mao et al. 2007; Sakasi et al. 2008).

Truncated protein products caused by deletions ranging from 1 to 7 bp were also observed in ASPM, and CENPJ genes in autosomal recessive primary microcephaly. Heterozygous subjects are asymptomatic, showing that short deletions might be often found, even in healthy subjects (Leal et al. 2003; Bond et al. 2003; Gul et al. 2006).

Screening genetics variations from ESTs public data sets, guided by expression studies, are a promising approach and

Table 1 Predicted deletions causing premature protein truncation

ESTs CODE	Gene position	MUTATION	PROTEIN SIZE
CD609928	COX-2/Exon5	c.[623delT] Frameshift Incorporating Ile162del	168aa
CB960307	COX-2/Exon5	c.[634delT] Frameshift Incorporating Leu166del	168aa
CD609931	COX-2/Exon5	c.[743delT] Frameshift Incorporating Pro203del	203aa
BX363285	COX-2/Exon7	c.[998delC+ 1018_1019 delCA + 1023_1024 delAG + 1045_1046delCA] Frameshift Incorporating Thr278del	291aa
AL712736	COX-2/Exon 9	c.[1412delT] Frameshift Incorporating Pro496del	513aa
BG121477	IL-1 β /Exon 6	c.[573delT] Frameshift Incorporating Phe162del	185aa
ES309163	IL-1 β /Exon 6	c.[630delG] Frameshift Incorporating Asn182del	185aa
BI819099	IL-1 β /Exon 6	c.[643delT] Frameshift Incorporating Ser186del	189aa
BE439432	IL-1 β /Exon 6	c.[573delT + 632delA] Frameshift Incorporating Phe162del	190aa
BE619081	IL-1 β /Exon 6	c.[612delC + 623_629delAGGAAAA] Frameshift Incorporating Leu176del	190aa
CA307234	IL-1 β /Exon 7	c.[732delU] Frameshift Incorporating Phe215del	220aa
BM999417	IL-1 β /Exon 7	c.[733delG + 1415_1416delUA] Frameshift Incorporating Val216del	220aa
BG252492	IL-1 β /Exon 7	c.[710delA + 857delC + 885delT + 970_977delAGGTTTTT] Frameshift Incorporating Lys208del	220aa
BE787882	IL-1 β /Exon 7	c.[857delC] Frameshift Incorporating Gln257del	259aa
BG186437	PLA2G4A/Exon 7	c.[639delG + 720delT] Frameshift Incorporating Arg154del	165aa
BG675386	PLA2G4A/Exon 11	c.[1300delC] Frameshift Incorporating Asp366delfs	389aa
AU120946	PLA2G4A/Exon 14	c.[1579delT + 1614delT] Frameshift Incorporating Ser458delfs	468aa
BE439441	PLA2G4A/Exon 15	c.[1829delA + 1836delA + 1869delA + 1907delG] Frameshift Incorporating Ser542delfs	571aa

demand further investigation in proper biological cyberinfrastructure that will optimize pipelines similar to the one proposed at this article (Stein 2008)

This will be a key initiative, in order to handle the overwhelming amount of data generated at our current "Omics Era" and will eventually maximize our understanding about the underlying molecular panel in neurodegeneration.

Acknowledgments We are greatly indebted with Aaron Rowe for reviewing this manuscript. This study was support by the following Brazilian funding agencies and academic bureaus: PROPESQ-UFPE, CAPES, CNPq, and FACEPE.

References

- Akiyama, H., Barger, S., Barnum, S., et al. (2000). Inflammation and Alzheimer's disease. *Neurobiology of Aging*, 21, 383–421. doi:10.1016/S0197-4580(00)00124-X.
- Barker, G., Batley, J., O'Sullivan, H., Edwards, K. J., & Edwards, D. (2003). Redundancy based detection of sequence polymorphisms in expressed sequence tag data using auto SNP. *Bioinformatics (Oxford, England)*, 19, 421–422. doi:10.1093/bioinformatics/btt881.
- Blaalock, E. M., Geddes, J. W., Chen, K. C., Porter, N. M., Markesbery, W. R., & Landfield, P. W. (2004). Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 2173–2178. doi:10.1073/pnas.0308512100.
- Bond, J., Scott, S., Hampshire, D. J., et al. (2003). Protein-truncating mutations in ASPM cause variable reduction in brain size. *American Journal of Human Genetics*, 73, 1170–1177. doi:10.1086/379085.
- Breloer, M., & Fleischer, B. (2008). CD83 regulates lymphocyte maturation, activation and homeostasis. *Trends in Immunology*, 29, 186–194. doi:10.1016/j.it.2008.01.009.
- Colangelo, V., Schurr, J., Ball, M. J., Pelaez, R. P., Bazan, N. G., & Lukiw, W. J. (2002). Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *Journal of Neuroscience Research*, 70, 462–473. doi:10.1002/jnr.10351.
- Coppola, G., Karydas, A., Rademakers, R., et al. (2008). Gene expression study on peripheral blood identifies progranulin mutations. *Annals of Neurology*, 64, 92–96. doi:10.1002/ana.21397.
- Fertuzinhos, S. M., Oliveira, J. R., Nishimura, A. L., et al. (2004). Analysis of IL-1alpha, IL-1beta, and IL-1RA [correction of IL-1RA] polymorphisms in dysthymia. *Journal of Molecular Neuroscience*, 22, 251–256. doi:10.1385/JMN:22:3:251.
- Forment, J., Gilbert, F., Robles, A., Conejero, V., Nuez, F., & Blanca, J. M. (2008). EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration. *BMC Bioinformatics*, 9, 5–15. doi:10.1186/1471-2105-9-5.
- Grimaldi, L. M. E., Casadei, C. M., Ferri, C., et al. (2000). Association of early-onset Alzheimer's disease with an interleukin-1α gene polymorphism. *Annals of Neurology*, 47, 361–365. doi:10.1002/1531-8249(200003)47:3<361::AID-ANA12>3.0.CO;2-N.
- Gul, A., Hassan, M. J., Hussain, S., Raza, S. I., Chishti, M. S., & Ahmad, W. (2006). A novel deletion mutation in CENPJ gene in a Pakistani family with autosomal recessive primary microcephaly. *Journal of Human Genetics*, 51, 760–764. doi:10.1007/s10038-006-0017-1.
- Heneka, M. T., & O'Banion, M. K. (2007). Inflammatory processes in Alzheimer's disease. *Journal of Neuroimmunology*, 184, 69–91. doi:10.1016/j.jneuroim.2006.11.017.
- Kidd, J. M., Cooper, G. M., Donahue, W. F., et al. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature*, 453, 56–64. doi:10.1038/nature06862.
- Leal, G. F., Roberts, E., Silva, E. O., Costa, S. M. R., Hampshire, D. J., & Woods, C. G. (2003). A novel locus for autosomal recessive primary microcephaly maps to 13q12.2. *Journal of Medical Genetics*, 40, 540–542. doi:10.1136/jmg.40.7.540.
- Mao, G., Pan, X., Zhu, B.-B., et al. (2007). Identification and characterization of OGG1 mutation in patients with Alzheimer's disease. *Nucleic Acids Research*, 35, 2759–2766. doi:10.1093/nar/gkm189.
- Minghetti, L. (2004). Cyclooxygenase-2 (COX-2) in inflammatory and degenerative brain diseases. *Journal of Neuropathology and Experimental Neurology*, 63, 901–910.
- Miller, J. A., Oldham, M. C., & Geschwind, D. H. (2008). A systems level analysis of transcriptional changes in Alzheimer's Disease and normal aging. *The Journal of Neuroscience*, 28, 1410–1420. doi:10.1523/JNEUROSCI.4098-07.2008.
- Navratil, V., Penel, S., Delmotte, S., Mouchiroud, D., Gautier, C., & Aouacheria, A. (2008). DigiPINS: a database for vertebrate exonic single nucleotide polymorphisms and its application to cancer association studies. *Biochimie*, 90, 563–569. doi:10.1016/j.biochi.2007.09.017.
- Nicoll, J. A. R., Mraz, R. E., Graham, D. I., et al. (2000). Association of interleukin-1 gene polymorphism with Alzheimer's disease. *Annals of Neurology*, 47, 365–368. doi:10.1002/1531-8249(200003)47:3<365::AID-ANA13>3.0.CO;2-G.
- Oliveira, J. R. M., Nishimura, A. L., Lemos, R. R., & Zatz, M. (2009). The genetics of disease in Brazil: 10 years of analysis in a unique population. *Journal of Molecular Neuroscience*, 37, 74–79. doi:10.1007/s12031-008-9124-0.
- Picoult-Newberg, L., Ideker, T. E., Pohl, M. G., et al. (1999). Mining SNPs from EST databases. *Genome Research*, 9, 167–174.
- Ray, S., Britschgi, M., Hebert, C., et al. (2007). Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling. *Nature Medicine*, 13, 1359–1362. doi:10.1038/nm1653.
- Sakasi, A., Horikawa, Y., Suwa, T., et al. (2008). Case report of familial carney complex due to novel frameshift mutation c.597delC (p.phe200LeufsX6) in PRKAR1A. *Molecular Genetics and Metabolism*, 95, 182–187. doi:10.1016/j.ymgme.2008.07.009.
- Stefansson, H., Rujescu, D., Cichon, S., et al. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature*, 455, 232–236. doi:10.1038/nature07229.
- Stein, L. D. (2008). Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. *Nature Reviews Genetics*, 9, 678–688. doi:10.1038/nrg2414.
- Stephenson, D. T., Lemere, C. A., Selkoe, D. J., & Clemens, J. A. (1996). Cytosolic phospholipase A2 (cPLA2) immunoreactivity is elevated in Alzheimer's disease brain. *Neurobiology of Disease*, 3, 51–63. doi:10.1006/mbdi.1996.0005.
- Tang, J., Vosman, B., Voorrips, R. E., Gerard van der Linden, G., & Leunissen, J. A. M. (2006). QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. *BMC Bioinformatics*, 7, 438–452. doi:10.1186/1471-2105-7-438.

- Tang, J., Leunissen, J. A. M., Voorrips, R. E., Gerard van der Linden, G., & Vosman, B. (2008). HaploSNPer: a web-based allele and SNP detection tool. *BMC Genetics*, 9, 23–29. doi:10.1186/1471-2156-9-23.
- Tuppo, E. E., & Arias, H. R. (2005). The role of inflammation in Alzheimer's disease. *The International Journal of Biochemistry & Cell Biology*, 37, 289–305. doi:10.1016/j.biocel.2004.07.009.
- Useche, F. J., Gao, G., Hanafey, M., & Rafalski, A. (2001). High-throughput identification, database storage and analysis of SNPs in EST sequences. *Genome Informatics*, 12, 194–203.
- Walsh, T., McClellan, J. M., McCarthy, S. E., et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, 320, 539–543. doi:10.1126/science.1155174.
- Wisniewski, T., & Konietzko, U. (2008). Amyloid-beta immunisation for Alzheimer's disease. *The Lancet Neurology*, 9, 805–811. doi:10.1016/S1474-4422(08)70170-4.
- Yu, C. E., Dawson, G., Munson, J., et al. (2002). Presence of large deletions in kindreds with autism. *American Journal of Human Genetics*, 71, 100–115. doi:10.1086/341291.

Table 1. Additional Polymorphisms in CD83 Gene

Location	Predicted nucleotide variation in coding and non coding sequences	Predicted protein variation in coding sequences. fs: frameshift	ESTs code
Exon 1	90delG (5' UTR)		BI 916827
	168_170delCGC (5' UTR)		BX361655
	164delC (5' UTR)		DB200908
Exon 2	c.256_258del TTG	Cys27del	DA584731
	c.268delT	Asp30de lfs	BP296648
	c.268delT	Asp30de lfs	BP296531
	c.299_300delCC	Pro41del fs	DB489168
	c.305delG	Val43del fs	AL552104
Exon 4	c.595_597delGAA	Lys140del	BI907532
	c.617delC	Leu147del fs	AU309536
	c.628delT	Leu151del fs	BE830187
	c.628delT	Leu151del fs	BE830134
Exon 5*	c.680delC	Leu168del fs and Stop Introduction in position 193aa	BG705577
	c.680delC+693delT	Leu168del fs and Stop Introduction in position175aa	BI758803
	c.747delC	Pro190del fs and Stop Introduction in position 193aa	BU192557
	906delG+926_933del/CATCTGGA	Region 3' UTR	AL575500
	1387delU+1392delU	Region 3' UTR	BE622865
	1544delA	Region 3' UTR	BG024615
	1443delT+1472delT+1489delT+1518delT+1704delT+1812delT	Region 3' UTR	BG177706
	1722delU+1729delG+1765delA+1813delT	Region 3' UTR	D51194
	1072delT	Region 3' UTR	BE622272
	1009delC+1029_1030delCA+1038delA	Region 3' UTR	R13302
	1061delT	Region 3' UTR	R19230
	1001delC+1024delC+1044delC	Region 3' UTR	BI762179
	968_969delGA	Region 3' UTR	CT004518
	841delG+860_863delTGGA+882delA+892delT+925delA	Region 3' UTR	BX390623
	872delG+994_997delATGA+1003_1005delCTG	Region 3' UTR	BM546803
	899delC+948_950delCAT	Region 3' UTR	BI769419
	826delA+902delT+942delT	Region 3' UTR	BM926451
	874delA+915delC	Region 3' UTR	CD641481
	847delC+880delC	Region 3' UTR	BE883894
	799_800delCA+803delT	Region 3' UTR	BQ231391
	780delA	Region 3' UTR	CV024063
	968delG+973delC+1058delC	Region 3' UTR	AL552104
	948delC	Region 3' UTR	BX361655
	943delC+945delC+972delA	Region 3' UTR	BI916827

Exon 5- Variant A*

Table 2. Additional Polymorphisms in COX-2, IL-1 α and PLA2G4A genes

Location	Predicted nucleotide variation in coding and non coding sequences	Predicted protein variation in coding sequences. fs: frameshift	ESTs code
Exon 1/COX-2	132C>G +133G>C (5' UTR)		DB019295; DA882455; DA277667; DN995052; AL583668; CN337323
Exon 3/ COX-2	c.[387delT + 440C>G]	Trp85del fs and Stop Introduction in position 88aa	CD609931 DN995052 DB480063 DB019295 DB223983 DA676889 DA170263 DA182079 DA882455 DA677048 DA670802 DA670627 DA486815 DA549596 DA994617 DA277667 DA674771 DA673460 DA676702 DA673400 DA675888 DA671756 DA674666 DA678032 DA675968 DA672971 DA672566 CV023684 CN337323 CD609928 CB960307 BX363285 BP337788 BP335244 BP337860 BP336843 BE785396 AL583668 BP335898 BE717711
Exon 3/ COX-2	c.440C>G	Val102Val Synonymous Mutation	CD609928
Exon 5/ COX-2	c.623delT	Ile162del fs and Stop Introduction in position 168aa	CD609928
Exon 5/ COX-2	c.634delT	Leu172del fs and Stop Introduction in position 168aa	CB960307
Exon 5/ COX-2	c.743delT	Pro203del fs	CD609931
Exon 7/ COX-2	c.[998delC +1018_1019 del CA+1023_1024 delAG+1045_1046delCA]	Thr278del fs and Stop Introduction in position 291aa	BX363285
Exon 9/COX-2	c.1412delT	Pro496del fs and Stop Introduction in position 513aa	AL712736
Exon 5/ IL-1A	c.[1308delT+1312delT+1326del G]	116-122 Phe,Ser,Phe,Leu,Ser,Asn,Val>Leu,Ala,Pro,Glu, Gln,Cys Lys122Lys Synonymous Mutation	BE788845
Exon 6/ IL-1A	c.1498delA	Ile180del fs and Stop Introduction in position 182aa	BM009964
Exon 1/PLA2G4A	88delA (5' UTR)		BP280758
Exon 2/PLA2G4A	174delG (5' UTR)		BG186437
Exon 7/PLA2G4A	c.[639del G +720del T]	Arg154del fs and Stop Introduction in position 165aa	BG186437
Exon 11/ PLA2G4A	c.1300delC	Asp366del fs and Stop Introduction in position 389aa	BG675386
Exon 14/ PLA2G4A	c.[1579delT+1614delT]	Ser458del fs and Stop Introduction in position 468aa	AU120946
Exon 15/PLA2G4A	c.[1829delA+1836delA+1869delA+1907delG]	Ser542del fs and Stop Introduction in position 571aa	BE439441
Exon 16/ PLA2G4A	c.2157A>G	Lys651Arg	CD104330; CB985872; T27656

Table 3. Additional Polymorphisms in IL -1 β

Location	Predicted nucleotide variation in coding and non coding sequences	Predicted protein variation in coding sequences. fs: frameshift	ESTs code
Exon 6	c.573delT	Phe162del fs and Stop Introduction in position 185aa	BG121477
	c.[573delT+632delA]	Phe162del fs and Stop Introduction in position 190aa	BE439432
	c.576delA	Gln164del fs	BP336652
	c.[612delC+623_629delAGGAAAA]	Leu176del fs and Stop Introduction in position 190aa	BE619081
	c.630delG	Asn182del fs and Stop Introduction in position 185aa	ES309163
	c.643delT	Ser186del fs and Stop Introduction in position 185aa	BI819099
	c.657delA	Asp 191del fs	BI907808
	c.657delA	Asp 191del fs	BG250625
Exon 7	c.[700_701delAA+710delA+752delA+839delG]	Asn205del + 206Tyr>Leu+ Lys209Lys Synonymous Mutation	AI609005
		+Lys213 del fs and Stop Introduction in position 220aa	
	c.700_701del AA	Asn205del+206Try>Leu fs and Stop Introduction in position 223aa	BI519707
	c.732delU	Phe215del fs and Stop Introduction in position 220aa	CA307234
	c.[733delG+1415_1416delUA]	Val216del fs and Stop Introduction in position 220aa	BM999417
	c.[710delA+857delC+ 885delT+970_977del AGGTTTTT]	Lys208del fs and Stop Introduction in position 220aa	BG252492
	c.857delC	Gln257del fs Stop Introduction in position 259aa	BE787882
	1005_1006delGT (3' UTR)		CD521315
	1162delC+1181delC+1200delC (3' UTR)		BI871339
	1188_1194delTTCTAAG+1216delC (3' UTR)		BM556327
	1129delC+1162delC+1181delC (3' UTR)		BG252897

7. ANEXO

RESULTADOS COMPLEMENTARES

A técnica de screening *in silico* de variações genéticas a partir de ESTs vem se tornando uma metodologia rápida e de baixo custo (PICOULT-NEWBERG *et al.*, 1999; USECHE *et al.*, 2001; NAVRATIL *et al.*, 2008; LEMOS *et al.*, 2009). Esse tipo de abordagem envolve uma grande quantidade de dados e podem indicar potenciais novos polimorfismos, que serão alvos de pesquisas de genotipagem experimental na população.

Os estudos de *microarrays* têm sido interpretados de diferentes maneiras. Como o resultado final pode ser visto como uma “foto” instantânea de um dado fenômeno biológico, alguns autores preferem se utilizar destes resultados apenas para gerar hipóteses, enquanto outros preferem inferir toda uma série de eventos metabólicos que possam ser explicados por uma suposta sequência de interações de genes que estavam *up-regulated* e *down regulated* (GESCHWIND *et al.*, 2000).

Neste estudo foi adotada a primeira posição, de utilizar os dados obtidos em experimentos anteriores para identificar genes candidatos relevantes para a predisposição para a DA. A identificação de variações genéticas em genes cuja expressão encontra-se alterada, nas lâminas de *microarrays* de pacientes com DA, pode vir a elucidar a natureza das mudanças fenotípicas e genotípicas dessa complexa doença.

Estes resultados complementares incluem a análise de três outros genes que estão associados com via inflamatória e um outro com a degradação das placas *B*-amilóide, estes se encontram listados na (Tabela 2).

Todos os genes foram provenientes de dois estudos de *microarray* de expressão de Colangelo *et al.* (2002) e de Weeraratna *et al.* (2007) onde amostras de tecido da região Cornu Ammonis 1 (CA1) do hipocampo e da região do Lóbulo Parietal Inferior (LPI) respectivamente, foram avaliadas em controles e pacientes com DA. As lâminas de *microarrays* para esses genes, revelaram expressão *up-regulated* e *down-regulated*.

Tabela 2. Lista dos nomes, localização e códigos de acesso dos quatro genes estudados.

Símbolo	Nomenclatura	Localização	RNAm
B94 ou TNFAIP2	Tumor necrosis factor, alpha-induced protein 2.	14q32	NM_006231
DPP1 ou CTSC isoforma A	Cathepsin C	11q14.1-q14.3	NM_001814
CXCR4 isoforma B	Chemokine (C-X-C motif) receptor 4	2q21	NM_003467
			NM_014693
ECE2 isoforma A/C	Endothelin converting enzyme 2	3	NM_032331

CA1 é uma subregião do hipocampo relacionada com a cognição e formação da memória, sendo neste local onde se inicia os depósitos de emaranhados neurofibrilares, tornando-se muito mais severo no decorrer da doença (KAUFMANN *et al.*, 1998. ROSSLER *et al.*, 2002).

O lóbulo parietal inferior é uma região que concentra funções sensoriais, integração sensorial motor, orientação espacial, e desenvolvimento da linguagem (CLOWER *et al.*, 2001).

A metodologia abordada foi de acordo com Lemos *et al.* (2009) com o acréscimo de duas etapas, a pesquisa sobre a associação de doenças e SNPs em regiões codificantes e o significado funcional de SNPs em regiões 3'UTR tanto para os genes do trabalho anterior (LEMOS *et al.*, 2009) e para os outros quatro genes (Tabela 2).

O programa SAP foi utilizado para predizer se as trocas de aminoácidos provocadas pelos SNPs apresentaram associação com doenças. A mutação predita e a sequência protéica de interesse foram colocadas em campos específicos na página principal do site e em seguida o programa realiza uma pesquisa no seu banco de dados que possui 686 características derivadas da sequência protéica (Figura 1).

Para identificar se as variações genéticas ocorridas nas regiões 3'UTR das ESTs foram em sítios específicos de complementaridade de miRNAs, trechos dessas ESTs foram alinhados através de um BLASTn com sequências de miRNA do banco de dados mirBASE (Figura 2).

Identificação *in silico* de Deleções e Transições.

O número total de ESTs alinhados com os 4 genes foram de 2597, sendo que 366 seqüências foram selecionadas pelos dois filtros *e-value* e identidade (Tabela 3).

Algumas das seqüências alinhadas com os genes são ESTs próprios do gene, revelando assim 100% de identidade. Estas seqüências foram descartadas da análise, por não serem alvo da pesquisa de polimorfismos.

A identificação *in silico* revelou Deleções e Transições detectadas em regiões codificantes e nas regiões 5' ou 3' UTR (Tabela 3).

Tabela 3. Resumo da análise *in silico* dos genes: TNFAIP2, ECE2 isoforma A e C, CTSC isoforma A, CXCR4 isoforma B.

ESTs alinhados	2597
ESTs selecionadas através dos dois filtros	366
Deleções	248
Transições	250
Sinônimos	05
Não Sinônimos	353
SNPs em UTRs	17

Deleções: maiores classes de mudanças detectadas.

As deleções foram classificadas como a maior classe de variação na análise dos quatros genes.

As transições detectadas pelo CLCbio, mostraram ser um dado já previamente descrito e anotado no dbSNP, demonstrando para esse caso que a metodologia proposta por este trabalho, chegou a um dado experimentalmente já comprovado previamente em amostras de DNA genômico.

Dentre as classes de variações as deleções também foram os dados mais expressivos e inéditos observados por Lemos *et al.* (2009).

Como as deleções foram os originais polimorfismos e também a maior classe de variação, foi essencial para a pesquisa, conhecer à sua natureza e as distinguir de artefatos do sequenciamento.

Muitas das seqüências de ESTs ditas “reads” que estão depositadas no banco de dados não dispõem de informações acerca da nota de qualidade do sequenciamento que as geraram, sendo isto um tipo de limitação, para todos aqueles que manuseiam esses dados e que muitas vezes precisam entrar em contato com os curadores desses bancos para maiores detalhes.

Neste trabalho também foi adotada a estratégia para minimizar erros na interpretação dessas deleções foi fazer uso do filtro *e-value* abaixo de 10^{-20} . Este índice mostrou ser apropriado para este fim, pois apontou depois do BLAST, apenas seqüências adequadas para a análise.

Recentemente, Kidd *et al.* (2008) identificaram experimentalmente as deleções como a maior classe de mudança nos genoma de oito indivíduos, localizando 1.695 sítios polimórficos, incluindo: (747) deleções, (724) inserções e (224) inversões.

Análise de genomas de pacientes com doenças de ordem neurológica e comportamentais tais como: esquizofrenia, autismo e retardo mental, tem revelado que as microdeleções são responsáveis por variações estruturais dos genes, causando prejuízo em vias do neurodesenvolvimento (YU *et al.*, 2002; STEFANSSON *et al.*, 2008; WALSH *et al.*, 2008).

Mutações sinônimas e não sinônimas.

Depois da tradução das seqüências alinhadas foi verificado um maior número de mutações não sinônimas em relação às mutações sinônimas (Tabela 3).

Mutações não sinônimas resultantes de algumas transições do tipo (c.2157A>G; Lys651Arg) gene PLA2G4A observada por Lemos *et al.* (2009) e (c.556C>T;

Thr153Ile) gene CTSC isoforma A (Tabela 6) identificada nesta análise, já haviam sido descritas (rs2307198) e (rs217086) respectivamente.

As duas mutações (Lys651Arg) e (Thr153Ile) foram posteriormente avaliadas pelo programa SAP, que inferiu escores de confiança de (0.889) e (0.608) respectivamente, sendo o intervalo de confiança variando de 0-1 com 1 apresentando alta associação com doença e 0 baixa associação.

Essas predições foram feitas apenas através das características bioquímicas da sequência protéica do gene em questão. Esses resultados servirão de indicadores de polimorfismos de risco associados a doença a serem testados futuramente.

Algumas deleções provocaram a perda de um aminoácido e não causaram erros na sequência protéica, estas foram observadas nas ESTs alinhadas com exon 3 gene TNFAIP2 (Tabela 5); exon 6 gene CTSC isoforma A (Tabela 6); exon 3 gene ECE2 isoforma C (Tabela 7).

Alguns estudos de associação caso-controle não apenas investigam a presença da variação genética, mas também se essa alteração repercute em algum significado funcional (PORIER *et al.*, 1998, 1999; BORRONI *et al.*, 2002; VEPSÄLÄINEN *et al.*, 2007; CORTINI *et al.*, 2008; COPPOLA *et al.*, 2008).

Em uma análise de resposta diferenciada ao tratamento com inibidores de acetilcolinesterase em pacientes com DA, foi constatado que 80% daqueles que não possuíam o alelo $\epsilon 4$ APOE mostravam melhora significativa após 30 semanas de tratamento, enquanto 60% dos pacientes que eram positivos para o alelo $\epsilon 4$ APOE apresentam declínio considerável (PORIER *et al.*, 1999; 1998; BORRONI *et al.*, 2002).

Em um estudo com amostras de pacientes com DA oriundos da Filândia, foi demonstrado que a variação (495 C>T) na região 3'UTR (rs4988514) no gene da

Somatostatina, quando associado a forma alelica APOE $\epsilon 4$, elevava o risco para desenvolvimento da doença (VEPSÄLÄINEN *et al.*, 2007).

Uma nova variação verificada no gene da Progranulina (g.100169 C>T), que resultou na mutação (p.Gly35Arg), foi significativamente associado com amostras de pacientes com DA, posteriores análises qualitativas e quantitativas revelaram que essa mutação não degradada nem diminui o transcrito de RNAm. No entanto, uma predição *in silico* feita pelo programa de software POLYPHEN indicou que essa mutação afeta a estrutura da proteína e que, nesse caso, outros mecanismos patogênicos podem estar contribuindo para doença (CORTINI *et al.*, 2008).

Tradução de seqüências: *Frameshift* e Stop códons.

Várias mutações não sinônimas presentes na região codificadora podem também provocar mudança no quadro de leitura, ocasionando em *Frameshift* e Stop_códons.

Frameshift são mutações que desorganizam o quadro de leitura devido à inserção ou deleção de uma base não múltipla do códon responsável pelo aminoácido (aa) em questão. Mutações no quadro de leitura podem provocar stop códons, que são paradas na tradução da seqüência de nucleotídeos, pela introdução do códon de terminação, gerando formas truncadas da proteína. Essas mutações podem prejudicar a formação e função dessa proteína. Erros na seqüência de aa, *in vivo*, são frequentemente degenerados.

A forma truncada da proteína Ubiquitina B (UBB⁺) em pacientes sem demência exemplifica que essa proteína pode estar associada à ausência de doença, no entanto ela também pode ser interpretada como um primeiro sinal da neuropatologia da DA (VAN LEEUWEN *et al.*, 1998).

Um estudo de imunohistoquímica demonstrou que as formas UBB⁺ e APP⁺ foram mais presentes em um grupo de pacientes que não apresentavam a princípio uma história de demência, mas exibiam depósitos de β -amilóide no neocortex e emaranhados

neurofibrilares no córtex entorrinal o que qualifica o diagnóstico para DA (KONISHI *et al.*, 2003).

As demais deleções identificadas neste trabalho envolveram de 1 até 10pb e a maioria delas quando presentes na região codificadora provocam *frameshift* e alguns *stop codóns* (Tabela 4).

Formas truncadas de proteínas são observadas em famílias com microcefalia primária autossômica recessiva (MCPH), uma desordem geneticamente heterogênea onde o afetado apresenta reduzido tamanho da circunferência da cabeça e retardo mental. Indivíduos heterozigotos são assintomáticos para a MCPH, até o momento seis loci MCPH1-MCPH6 já foram identificados (LEAL *et al.*, 2003).

Deleções que variaram de 1 a 7pb provocaram *frameshifts* e formação de uma prematura proteína truncada, foram observadas no gene ASPM no locus MCPH5 e gene CENPJ locus MCPH6 (BOND *et al.*, 2003; GUL *et al.*, 2006; NICHOLAS *et al.*, 2008). Esses achados demonstram que apesar dos heterozigotos possuírem formas truncadas de proteínas, eles não manifestam a doença.

Clusters de variações genéticas foram identificados em três dos quatro genes estudados sendo eles: gene **TNFAIP2** *clusters* nos exons 8, 9 e 10 (Figura 3), gene **CTSC** isoforma A *clusters* no exon 5 um dos maiores, e exons 6 e 7 (Figura 3), gene **ECE2** isoforma C *clusters* no exon 3 (Figura 3). Essas regiões poderão ser alvos de novas variações candidatos a serem testados no futuro.

Variações nas regiões UTRs

As deleções foram detectadas nas regiões 5' e 3'UTR nos seguintes exon 1 e 11 gene TNFAIP2 (Tabela 5) e exon 22 gene ECE2 isoforma A e exon 3 gene ECE2 isoforma C (Tabela 7).

Foram feitos alinhamentos através do programa miRBASE entre trechos dos exons e ESTs que alinharam com os genes do nosso trabalho anterior (LEMOS *et al.*, 2009) e mais estes quatro genes. Foi verificada que duas seqüências de ESTs alinhadas com o exon 5a do gene CD83, apresentam similaridade com a seqüência de microRNA (hsa-miR-939) do gene humano, essas ESTs apresentam deleções ver (Figura 4). Esses resultados preliminares não indicam que essas deleções possam influenciar na perda de funções desses sítios.

Em uma primeira análise sistemática foi avaliado se potenciais sítios de mutações nas regiões 3'UTR de 79 genes associados a doenças conhecidas localizavam-se dentro de seqüências de microRNAs. Os alinhamentos mostraram que alguns SNPs alinharam com os sítios de microRNA, mas sem significado biológico que pudesse prejudicar a interação e funcionamento desses sítios (CHUZHANOVA *et al.*, 2007).

Outros SNPs tem de fato afetado a ligação do miRNA e RNAm, de 120.000 conhecidos SNPs na região 3'UTR, 19.913 provavelmente modificam essa interação destruindo alvos de sítios preservados (785) ou alvos não preservados (9470), ou mesmo criando novos alvos (10.283) (www.patrocles.org).

Análise do percentual GC

A análise geral do conteúdo GC dos exons que apresentaram deleções de cada gene estudado tanto os do nosso trabalho anterior (LEMOS *et al.*, 2009) quanto os quatro últimos genes podem ser visualizadas nas (Figuras 5,6,7,8,9,10,11,12,13,14).

No gene CD83 foi observado que o conteúdo GC teve distribuição em torno de 70% no exon 4 e 5a (Figura 5). No gene COX2 o exon 7 foi dentre os visualizados aquele que apresentou concentrados blocos de conteúdo GC em torno de 70% em relação aos exons 3, 5, e 9 (Figura 6). A distribuição de GC foi percentualmente maior

70% no exon 6 em relação ao exon 7 no gene IL-1 α (Figura 7). Equivalentes blocos de GC em torno de 70% foram visualizados nos exons 6 e 7 no gene IL-1 β (Figura 8).

Os exons 1, 2, e 15 apresentaram maior percentual GC cerca de 70% ao longo de todo exon quando comparado ao exon 17 no gene ECE2 isoforma A (Figura 9). Apenas exon 3 foi analisado no gene ECE2 isoforma C o exon 2b no gene CXCR4 isoforma B, ambas imagens demonstraram distribuição GC em torno de 70% (Figura 10 e 11).

Análise do conteúdo GC dos exons 2, 3, 4, 5, 7, 8, 9, 10, 11 do gene TNFAIP2 revelou através desses exons que esse gene apresenta de modo geral percentual GC em torno de 70% (Figura 12).

O conteúdo GC dos exons 7,11,14,15,16 do gene PLA2G4A apresentou percentuais em torno de 70%, com intervalos de com áreas de menor percentual (Figura 13). Os exons 6,7,9 gene CTSC apresentaram percentuais em torno de 50% de GC (Figura 14).

Essas indicações de percentuais de conteúdo GC podem distinguir deleções de artefatos do sequenciamento, uma vez que áreas ricas em GC, além de serem regiões ricas em genes, são também áreas que geralmente induzem erros de sequenciamento, devido às três pontes de hidrogênio que ligam esses nucleotídeos e dificultam a abertura da cadeia para a leitura das bases.

Este procedimento é mais uma ferramenta dessa metodologia para tentar apontar previamente a natureza das deleções detectadas *in silico* e se fazer ajustes na técnica do sequenciamento para os futuros trabalhos. Somente os experimentos de genotipagem poderão confirmar esses achados.

Genes *up-regulated* e *Down-regulated* envolvidos com a via inflamatória e degradação das placas *B*-amilóide.

A geração e acúmulo dos resíduos ($A\beta_{40-42}$) e dos emaranhados neurofibrilares são os iniciais mecanismos envolvidos com o processo patogênico da DA. Evidências recentes sugerem que o processo inflamatório seria o terceiro mecanismo envolvido, que uma vez iniciado pela degeneração, poderia contribuir para a progressão da doença (HENEKA; O'BANION, 2007; AKIYAMA *et al.*, 2000).

O processo inflamatório é apontado como o fator de aceleração para a progressão da doença, no entanto a resposta imune apresenta benefícios quando pode limitar a doença.

Polimorfismos genéticos, fatores epigenéticos (presença de outras doenças), infecções microbianas, injúrias traumáticas e uso de drogas, podem ser responsáveis pela transição de uma resposta inflamatória benéfica para um processo acelerado de neuroinflamação, além disso respostas diferenciadas podem estar associadas a diferentes níveis de depósitos placas *B*-amilóide, de emaranhados neurofibrilares e a presença ou ausência do alelo $\epsilon 4$ APOE (IQBAL, 2005).

O aumento na taxa de expressão dos genes CTSC e TNFAIP2 são associados à alta atividade dos processos inflamatórios e morte celular que acontece moderadamente no hipocampo de pacientes com DA (COLANGELO *et al.*, 2002).

Outro gene que também apresentou nível elevado da taxa de expressão em pacientes com DA foi o CXCR4, este também faz parte da via inflamatória. O gene ECE2 revelou diminuição de expressão, sendo este responsável pela degradação das placas *B*-amilóide, estes dois genes foram provenientes do tecido do lóbulo parietal inferior (WEERARATNA *et al.*, 2007).

Uma das classes de mediadores da inflamação são as citocinas moléculas envolvidas na resposta imune, que tanto podem promover a resposta inflamatória como causar a morte celular, sendo elas secretadas pela microglia e astrocitos a cerca das placas amilóides (STEINMAN, 2008).

Uma citocina envolvida na resposta inflamatória é o Fator de Necrose Tumoral alpha (TNF- α) classificada multifuncional, pois regula a reação inflamatória através da ativação de outros genes, sendo também capaz de induzir a morte celular por apoptose e a sua super expressão está associada à diversas doenças. (IMAIZUMIN *et al.*, 2008).

O gene TNFAIP2 é uma citocina induzida de início imediato pelo gene TNF- α em células endoteliais (RUSINIAK *et al.*, 2000). Taxa de expressão aumentada de TNFAIP2 foram atribuídos à ativação de genes inflamatórios como o TNF- α e IL- β no desencadeamento da resposta inflamatória (SARMA *et al.*, 1992; COLANGELO *et al.*, 2002).

Quimiocinas são também citocinas envolvidas com a motilidade celular, manutenção da arquitetura tecidual e resposta inflamatória. Receptores de quimiocinas são pequenas moléculas secretadas pelos Linfócitos T e B. A ligação da quimiocina e receptores servem para orientar os linfócitos para específicas regiões dentro dos linfonodos, para então ocorrer a resposta inflamatória (ABBAS, 2005).

O gene CXCR4 é um receptor de quimiocina expresso pelas células B, que atua no direcionamento dessas células para dentro dos folículos dos linfonodos, onde ocorrerá a reação inflamatória (ABBAS, 2005). Os receptores de quimiocinas são também expressos nas células do SNC, e a taxa de expressão aumentada de muitos desses receptores estão associadas aos tecidos envolvidos com a DA (ALBRIGHT *et al.*, 1999; XIA; HYMAN, 1999 WEERARATNA *et al.*, 2007).

Um outro gene envolvido com a resposta de células imune/inflamatória é o

CTSC, que codifica uma enzima denominada de Catepsina C, uma cisteína protease que age nos lisossomos e degradando polipéptidos (www.ncbi.nlm.nih.gov/entrez)

Os três genes com níveis elevados de expressão fazem parte de um conjunto de respostas inflamatórias ao tecido inflamado, cada um agindo num determinado momento e às vezes fazendo parte de uma cascata de reação.

O único gene que apresentou taxa reduzida de expressão em pacientes com DA, foi o ECE2, uma enzima responsável pela degradação das placas *B*-amiloide. Esse achado não a torna responsável pelo surgimento da doença, mas seu nível diminuído de expressão pode estar envolvido com a progressão DA, uma vez que sua função está comprometida (ECKMAN *et al.*, 2001, 2003; WEERARATNA *et al.*, 2007).

Genes associados com a via inflamatória apresentam um aumento na taxa de expressão em amostras de idosos sem demência, no entanto em pacientes com DA essa taxa apresenta um elevado acréscimo (BLALOCK *et al.*, 2004).

Estudos de microarrays de expressão mostram alterações nas taxas de expressão de genes envolvidos com as vias imune e inflamatória na DA, mas ainda não são capazes de responder se essas mudanças ocorrem no início e podem promover a doença ou são reações adversas do processo degenerativo (WYSS-CORAY *et al.*, 2006)

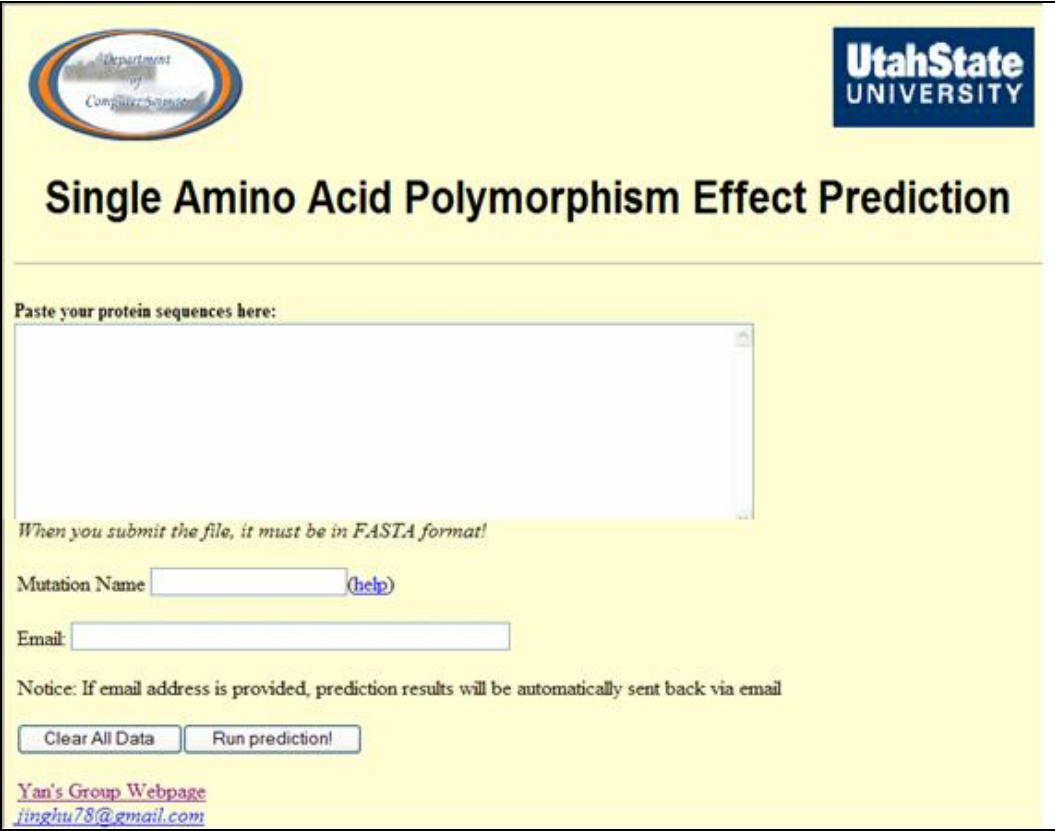
8. Conclusão

Após utilizar uma sequência de etapas associadas a diferentes ferramentas de bioinformáticas, foi possível verificar que o estudo de sequências de ESTs para se identificar novas variações candidatas é uma abordagem promissora e acessível.

Isto fica mais evidente se consideramos que a infra-estrutura necessária principal foi o uso de um computador ligado à internet e a associação de um *software* que é temporariamente disponível gratuitamente.

Além disso, pudemos validar a técnica ao mostrar que a mesma foi capaz de achar SNPs previamente descritos, além de destacar novos candidatos que serão posteriormente investigados em amostras de DNA de nossos estudos anteriores.

Estes achados levantam a possibilidade de que deleções de até 10 pares de bases estejam ligadas à fisiopatologia da DA, eventualmente em genes do sistema imunológico.



The screenshot shows a web application titled "Single Amino Acid Polymorphism Effect Prediction". At the top left is a logo for the "Department of Computer Science" and at the top right is the "Utah State UNIVERSITY" logo. The main heading is "Single Amino Acid Polymorphism Effect Prediction". Below this, there is a text input area labeled "Paste your protein sequences here:". A note below the input area states "When you submit the file, it must be in FASTA format!". There is a "Mutation Name" input field with a "(help)" link next to it. Below that is an "Email:" input field. A notice states "Notice: If email address is provided, prediction results will be automatically sent back via email". At the bottom, there are two buttons: "Clear All Data" and "Run prediction!". At the very bottom, there is a link to "Yan's Group Webpage" and an email address "jinghu78@gmail.com".

Department of Computer Science

Utah State UNIVERSITY

Single Amino Acid Polymorphism Effect Prediction

Paste your protein sequences here:

When you submit the file, it must be in FASTA format!

Mutation Name ([help](#))

Email:

Notice: If email address is provided, prediction results will be automatically sent back via email

[Yan's Group Webpage](#)
jinghu78@gmail.com

Figura 1. O programa SAP classifica a alteração protéica, originada pelos SNP, em uma mutação associada à doença ou mutação neutra, usando apenas a informação derivada da seqüência protéica.

Sequences
miRBase

miRBase::Sequences

MANCHESTER 1824
Wellcome Trust
Sanger Institute

Home Search Browse Genomics Help Download Submit miRBase Search

Search miRBase::Sequences

By sequence

Single sequence searches:
Paste a sequence here to search against miRNA sequences. You can choose to search against the intact precursor sequences or just the mature miRNAs. This search may take a few minutes. (Max size 1000bp's)

Search sequences: Mature miRNAs

Search method: BLASTN

Choose BLASTN to search for a miRNA in a longer sequence. SSEARCH is useful for finding a short sequence within the library of miRNAs (for instance, find a short motif in a miRNA or precursor stem-loop, or find mature sequences that are related to your query).

Or: Select the sequence file you wish to use

Procurar...

Figura 2. O programa miRBASE é uma ferramenta de alinhamento de trechos de seqüências de RNA contra um repositório de seqüências de microRNA.

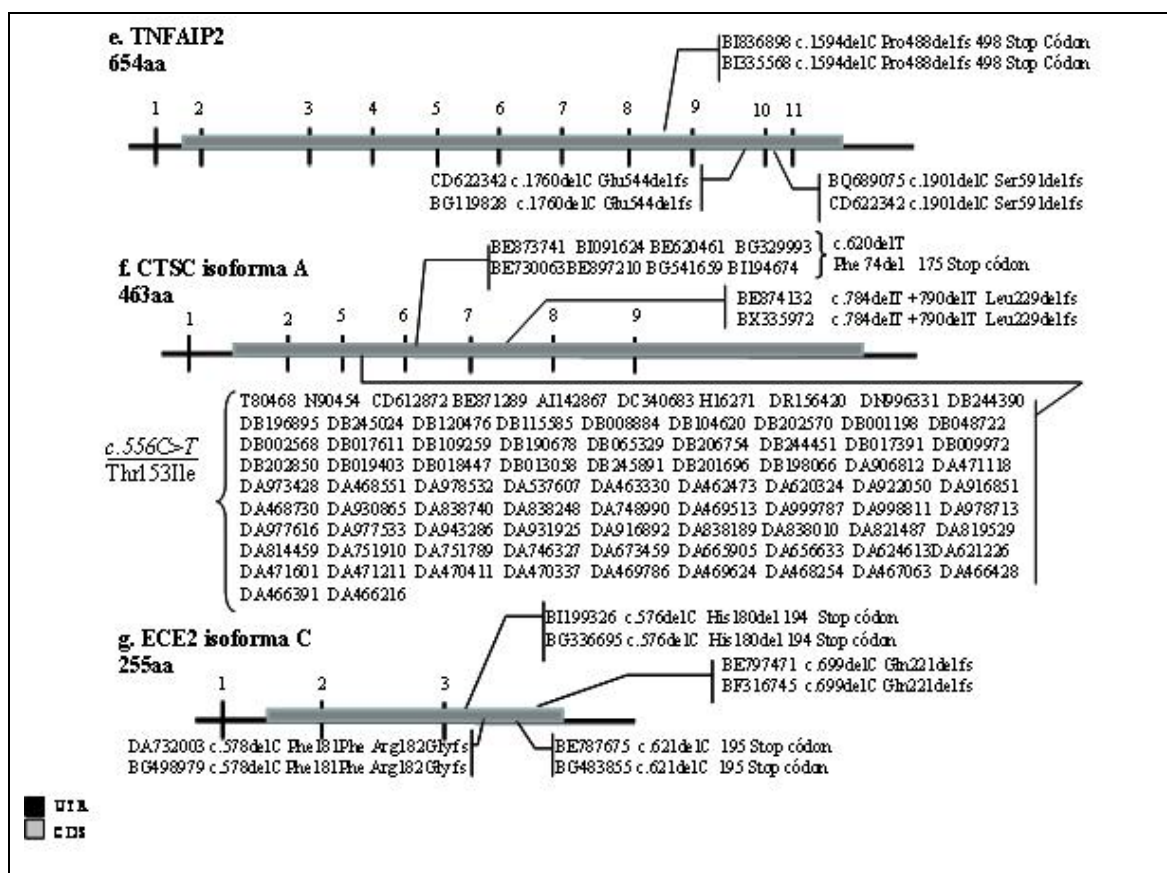


Figura 3. Identificação dos códigos de ESTs e das posições na região UTR e CDs, a parte sublinhada representa a variação previamente identificada nos gene CTSC isoforma A durante o processo de validação virtual.

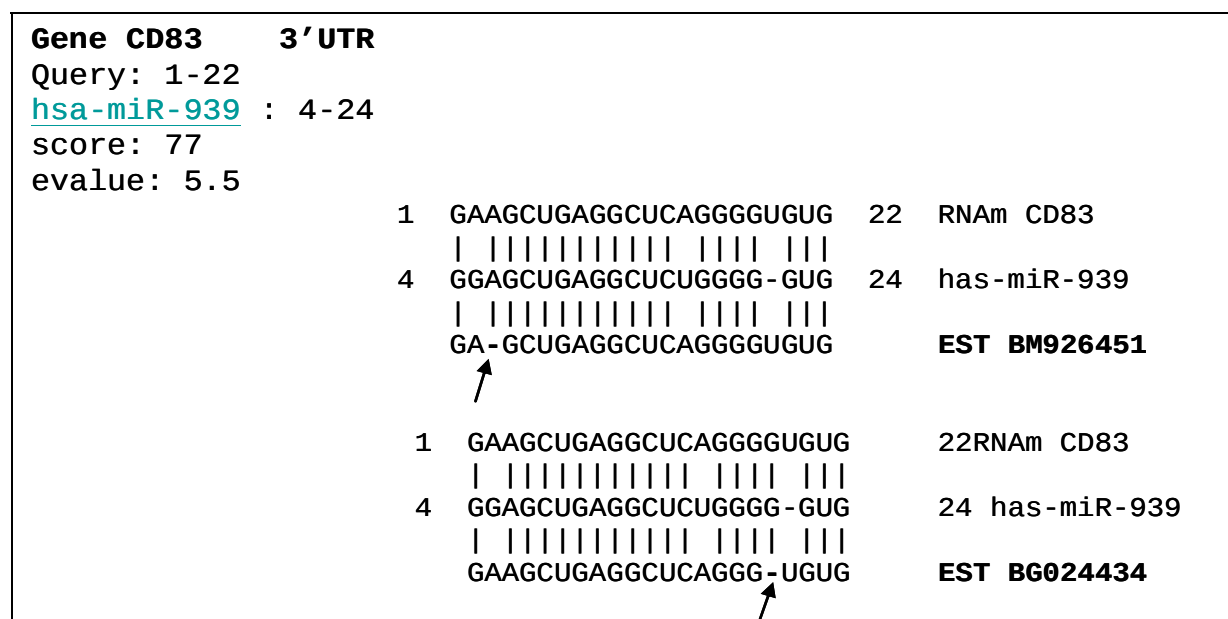


Figura 4. Alinhamento de duas ESTs (BM926451) e (BG024434) com o exon 5a e o miRNA (has-miR-939). A seta indica a posição da deleção.

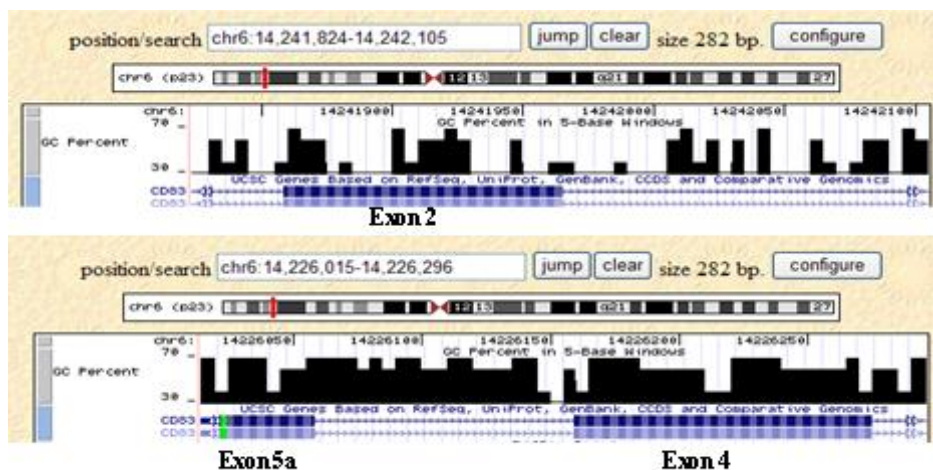


Figura 5. Análise do conteúdo GC dos exons 2, 4 e 5a do gene CD83 isoforma a.

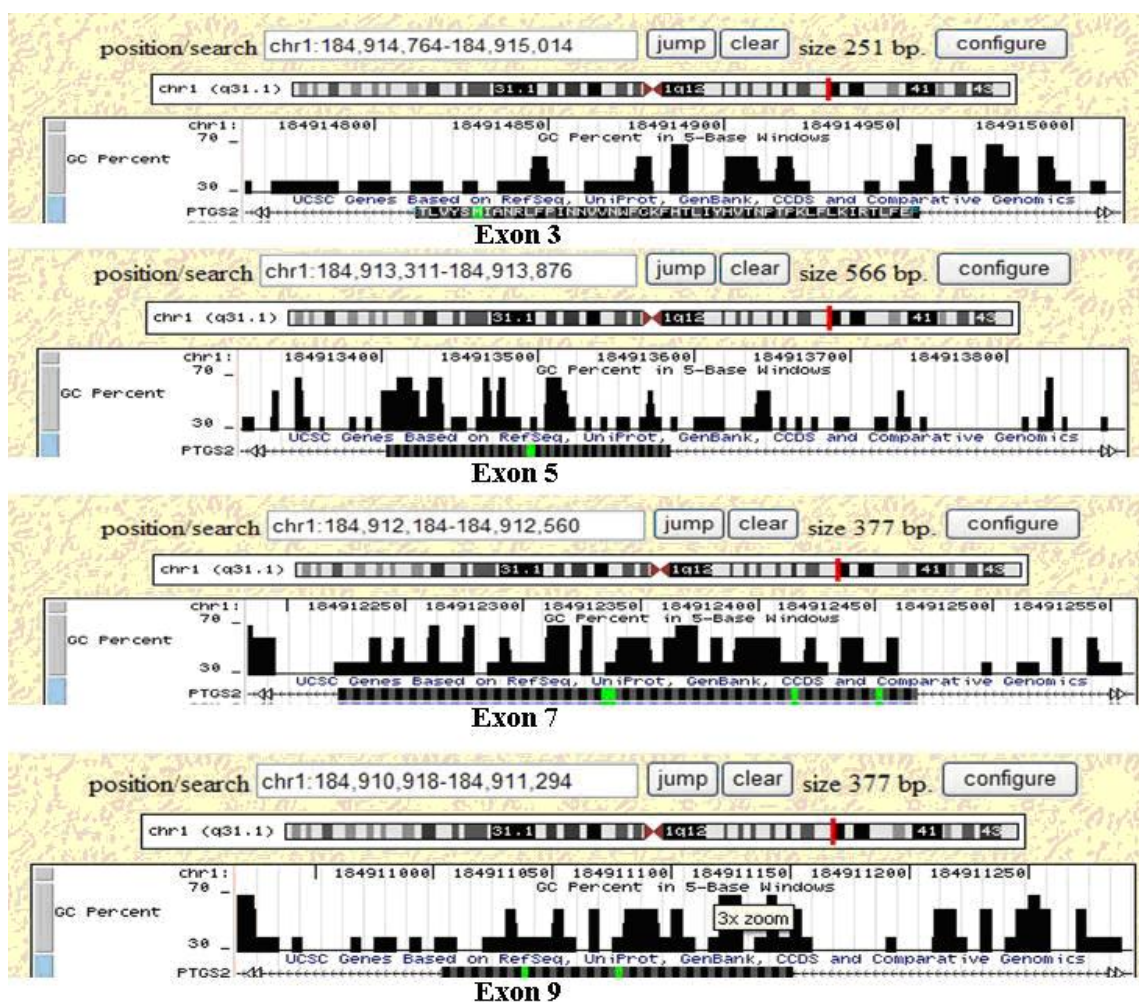


Figura 6. Análise do conteúdo GC dos exons 3, 5, 7 e 9 do gene COX2

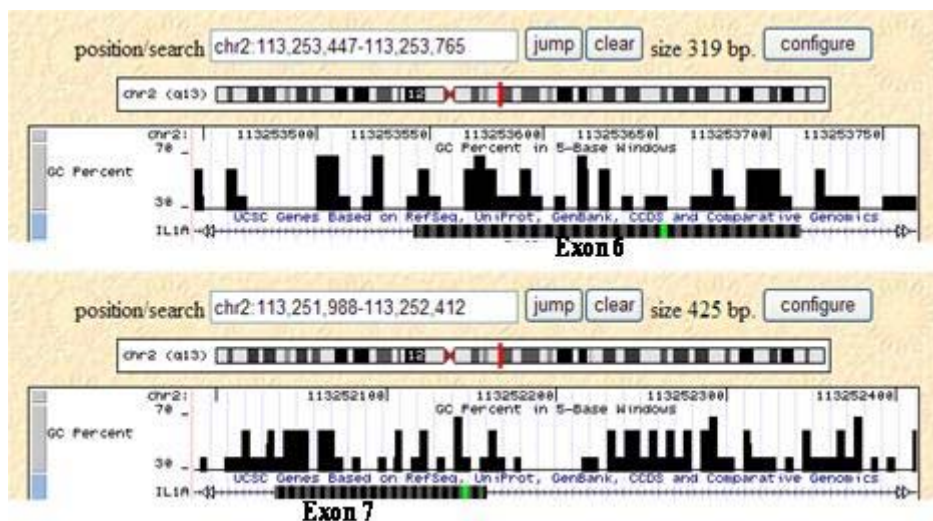


Figura 7. Análise do conteúdo GC dos exons 6 e 7 do gene IL-1 α .

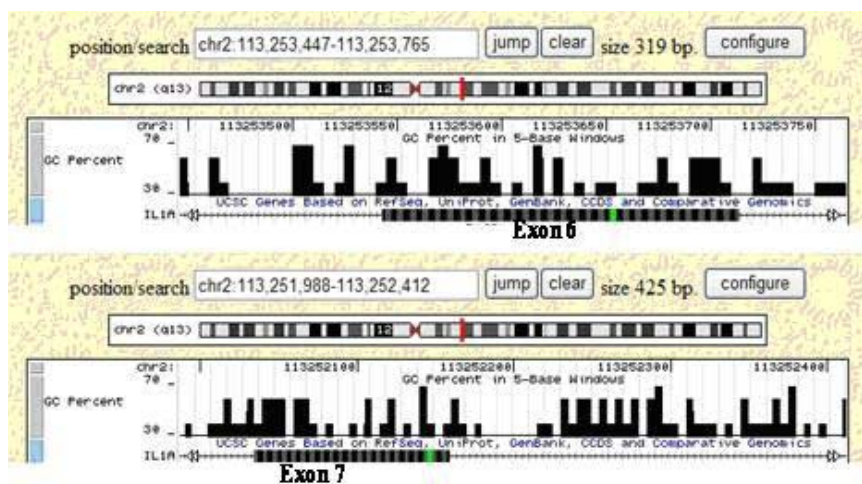


Figura 8. Análise do conteúdo GC dos exons 6 e 7 do gene IL-1 β

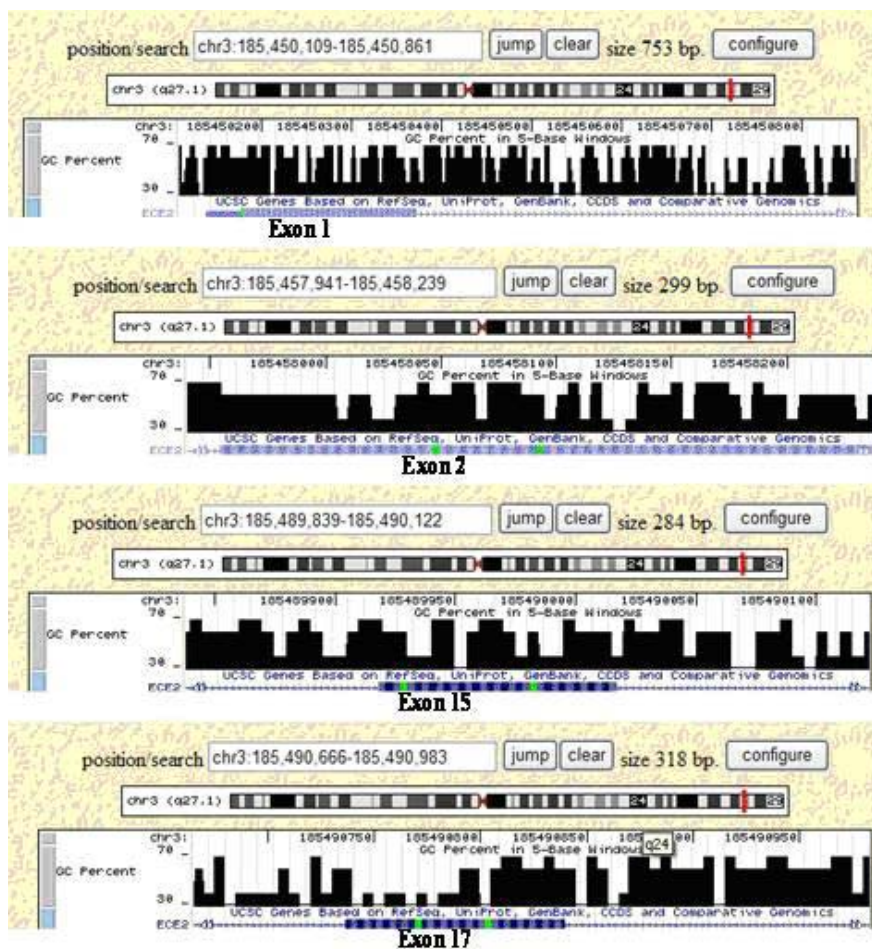


Figura 9. Análise do conteúdo GC dos exons 1, 2, 15 e 17 do gene ECE isoforma A.

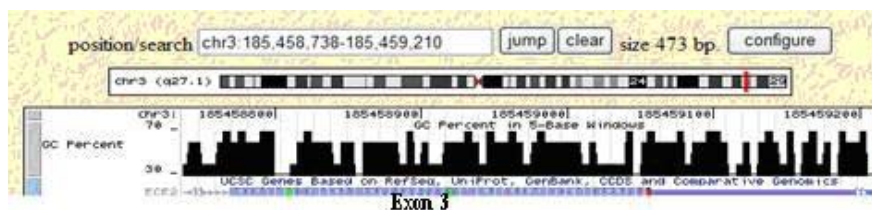


Figura 10. Análise do conteúdo GC dos exon 3 do gene ECE isoforma C.

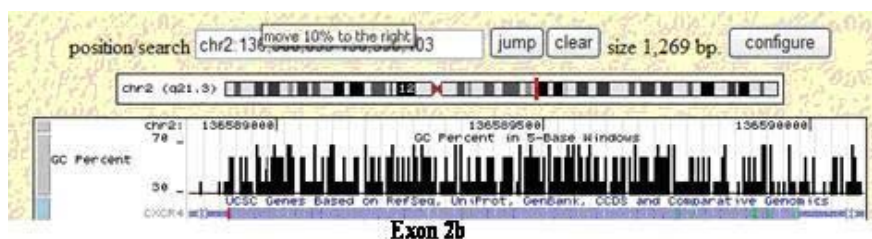


Figura 11. Análise do conteúdo GC dos exon 2b do gene CXCR4 isoforma B.

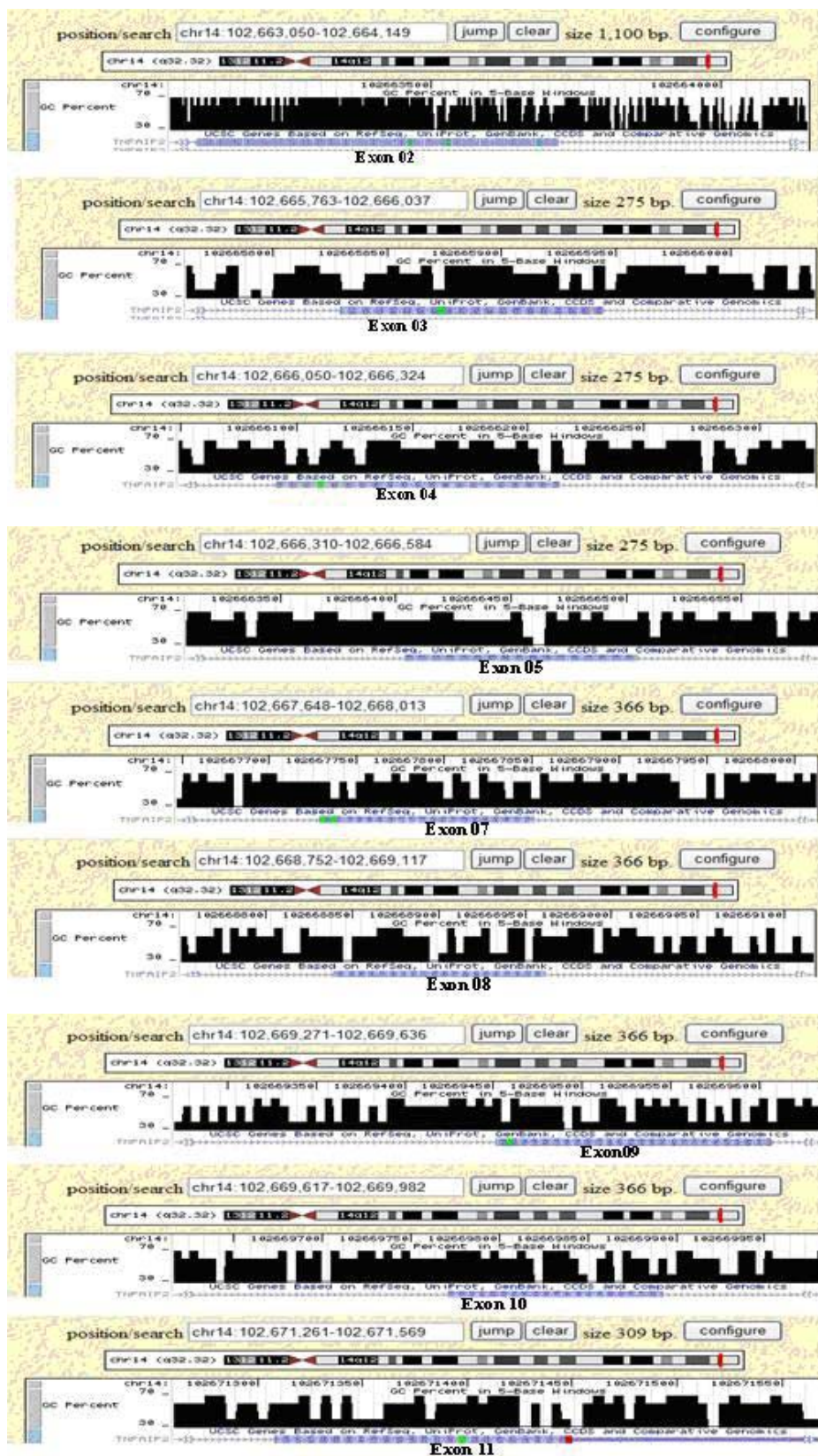


Figura 12. Análise do conteúdo GC dos exons 2, 3, 4, 5, 7, 8, 9, 10, 11 do gene TNFAIP2.

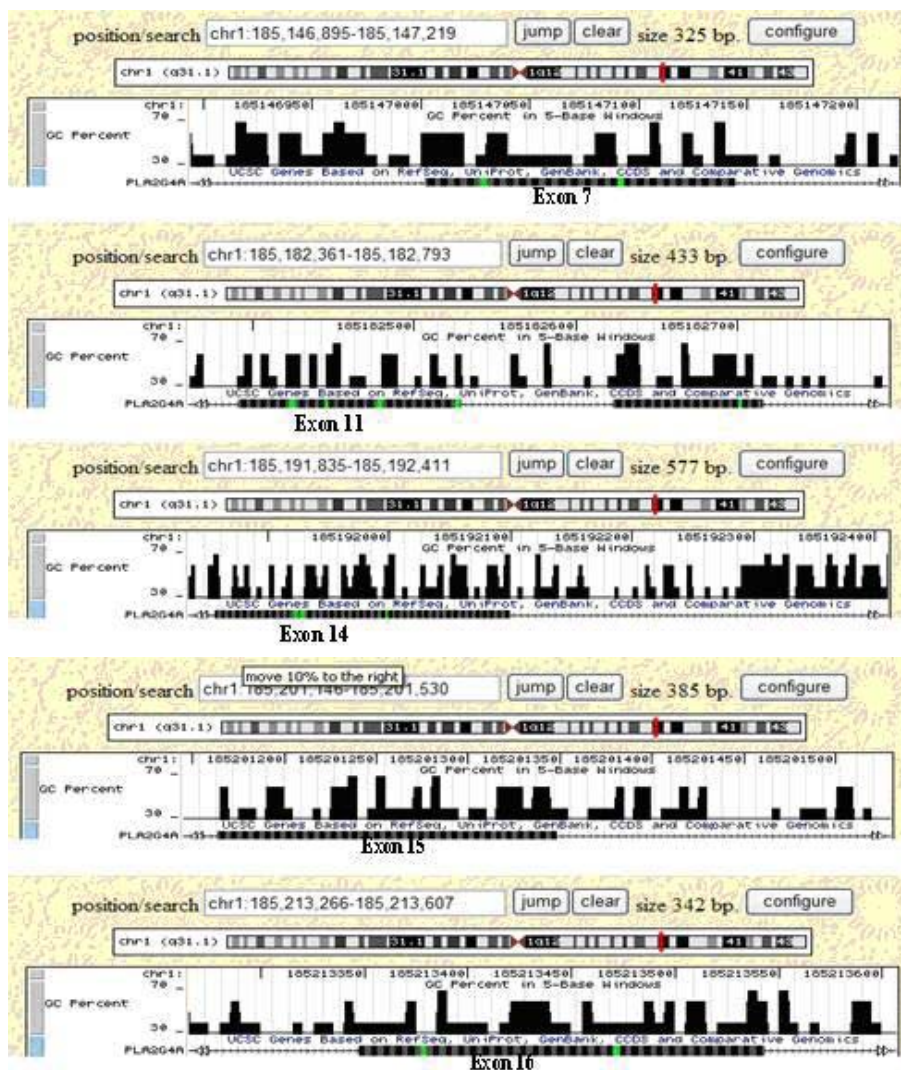


Figura 13. Análise do conteúdo GC dos exons 7, 11, 14, 15, 16 do gene PLA2G4A.

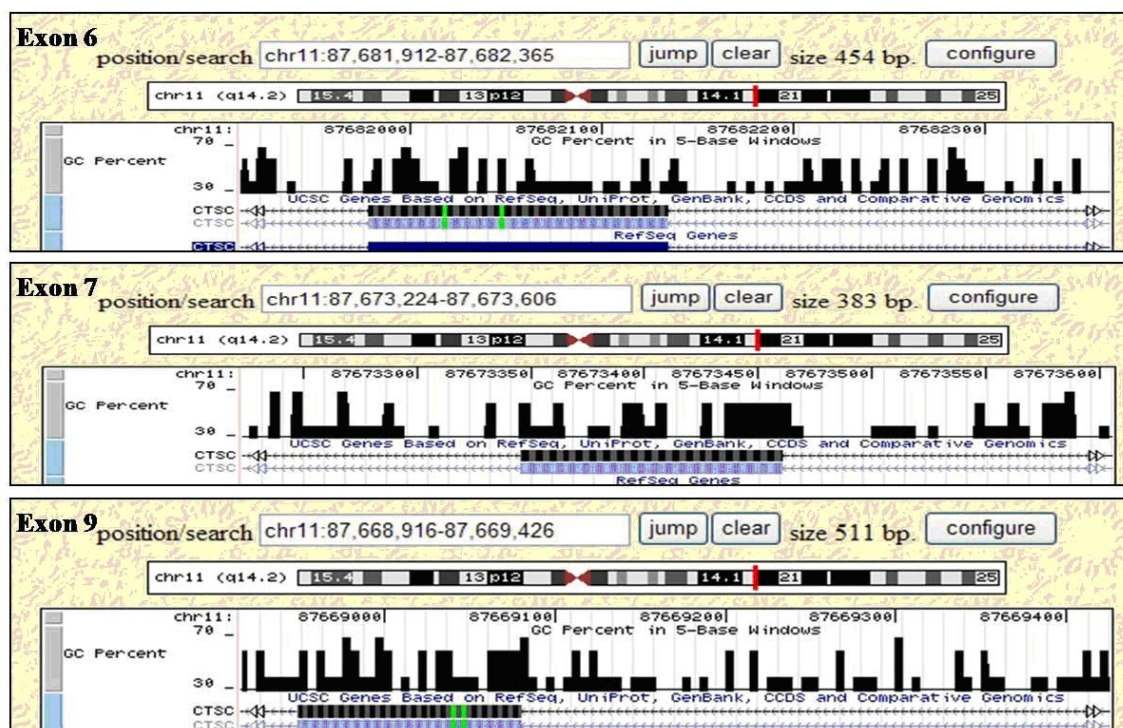


Figura 14. Análise do conteúdo GC dos exons 6, 7, 9 do gene CTSC.

Tabela 4. Deleções preditas que causam prematuras proteínas truncadas.

Código ESTs	Posição do gene	Variação	Tamanho da proteína
BF035953	ECE Isoforma A Exon 2	c.[452_457delTGGGGA] Gly Glu 139_140del Incorporação de Frameshift	141aa
BU568980	ECE Isoforma C Exon 3	c.[618delT] Ser194Pro Incorporação de Frameshift	195aa
AA844606	TNFAIP2 / Exon 4	c.[1134delG] Arg335del Incorporação de Frameshift	338aa
BG831039	TNFAIP2 / Exon 7	c.1468delC c.1695delC	446aa
AA464440	TNFAIP2 / Exon 9	His522del Incorporação de Frameshift c.658delC	525aa
BI828133	CTSC / Exon 6	Ala187del Incorporação de Frameshift c.654_655delAC	200aa
BI520194	CTSC / Exon 6	Thr186del Incorporação de Frameshift c.661delC	192aa
BE789469	CTSC / Exon 6	Thr188del Incorporação de Frameshift c.661delC+690delC	200aa
BG430120	CTSC / Exon 6	Thr188del Incorporação de Frameshift c.617delC	201aa
BG433205	CTSC / Exon 6	Phe174del Incorporação de Frameshift c.617delC	175aa
BE870710	CTSC / Exon 6	Phe174del Incorporação de Frameshift c.658delC +677_683delTGGAC	175aa
BQ224700	CTSC / Exon 6	Ala187del Incorporação de Frameshift c.651delT	201aa
BF130367	CTSC / Exon 6	Thr185del Incorporação de Frameshift c.617delC	200aa
BG432521	CTSC / Exon 6	Phe174del Incorporação de Frameshift c.926delA	175aa
CD612875	CTSC / Exon 8	Asn276del Incorporação de Frameshift c.882delU	283aa
BE785436	CTSC / Exon 8	Phe261del Incorporação de Frameshift c.882delU	266aa
BE874132	CTSC / Exon 8	Phe261del Incorporação de Frameshift c.901delA	266aa
AL569316	CTSC / Exon 8	Arg270del Incorporação de Frameshift c.931delC	237aa
AL542222	CTSC / Exon 8	Pro281del Incorporação de Frameshift c.931delC	283aa
BX399933	CTSC / Exon 8	Pro281del Incorporação de Frameshift c.931delC	283aa
BG542865	CTSC / Exon 8	Pro281del Incorporação de Frameshift c.924_925delCA + 931_932delCA	283aa
BX357941	CTSC / Exon 8	Asn276del Incorporação de Frameshift c.925_931delAACAATT	283aa
BE542347	CTSC / Exon 8	Asn276del Incorporação de Frameshift	283aa

Tabela 5. Adicionais polimorfismo no gene TNFAIP2.

Localização	Variações preditas de nucleotídeos em seqüências codificantes e não codificantes	Variações preditas de aminoácidos. fs: frameshift	ESTs
Exon 1	79delU (3'UTR)		W33183
Exon 2	c.[851delC + 870delG]	Ala1241 fs	AA854458
	c.[919_920CC>TG + 922_925delTGGC]	Ala>Val263+V264del fs	CD622344
Exon 3	c.1050delC	Leu1307 fs	AA772252
	c.1064delG	Arg311 del fs	AW628453
	c.1072_1074delGAC	Arg314del	BI914269
Exon 4	c.1134delG	Arg335del fs and Stop Introduction in position 338aa	AA844606
	c.[1179_1180delCA+1200delC]	Gln350del fs	AA772250
Exon 5	c.1250delC	Lys374del fs	AA772252
Exon 7	c.1449delT	Trp440del fs	BG119828
	c.1468delC	Stop Introduction in position 446aa	BG831039
	c.1517delT	Phe462del fs	BI836898
Exon 8	c.1594delC	Pro488del fs and Stop Introduction in position 498	BI836898
	c.1594delC	Pro488del fs and Stop Introduction in position 498	BI335568
	c.1664delC	Cys512del fs	BG685187
Exon 9	c.1743delC	Leu538del fs	BF094358
	c.1760delC	Glu544del fs	CD622342
	c.1760delC	Glu544del fs	BG119828
	c.1695delC	His522del fs and Stop Introduction in position 525aa	AA464440
	c.1763_1766delGCAG	Glu544del fs	BF869291
	c.1790delC	Leu553del fs	BE272545
Exon 10	c.1858delC	Pro576del fs	AA464440
	c.1901delC	Ser591del fs	BQ689075
	c.1901delC	Ser591del fs	CD622342
	c.1902_1903delCA	Ser591del fs	BM048080
Exon 11	2550delG 3'UTR		AI923266
	2698_2702delGCCCC (3'UTR)		AI923340
	2854delA (3'UTR)		AI923367
	2898delC (3'UTR)		AI923255
	3191delT (3'UTR)		AI923339

Tabela 6. Adicionais polimorfismo no gene CTSC.

Localização	Variações preditas de nucleotídeos em seqüências codificantes e não-codificantes	Variações preditas de aminoácidos Fs: frameshift	EST
Exon 5	c.556C>T	Thr153Ile	T80468 N90454 CD612872 BE871289 AI142867 DC340683 H16271 DR156420 DN996331 DB244390 DB196895 DB245024 DB120476 DB115585 DB008884 DB104620 DB202570 DB001198 DB048722 DB002568 DB017611 DB109259 DB190678 DB065329 DB206754 DB244451 DB017391 DB009972 DB202850 DB019403 DB018447 DB013058 DB245891 DB201696 DB198066 DA906812 DA471118 DA973428 DA468551 DA978532 DA537607 DA463330 DA462473 DA620324 DA922050 DA916851 DA468730 DA930865 DA838740 DA838248 DA748990 DA469513 DA999787 DA998811 DA978713 DA977616 DA977533 DA943286 DA931925 DA916892 DA838189 DA838010 DA821487 DA819529 DA814459 DA751910 DA751789 DA746327 DA673459 DA665905 DA656633 DA624613DA621226 DA471601 DA471211 DA470411 DA470337 DA469786 DA469624 DA468254 DA467063 DA466428 DA466391 DA466216
Exon 6	c.620delT	Phe174del e Stop introduzido na posição 175aa	BE873741 BI091624 BE620461 BG329993 BE730063 BE897210 BG541659 BI194674
Exon 6	c.658delC	Ala187del fs e Stop introduzido na posição 200aa	BI828133
Exon 6	c.654_655delAC	Thr186del fs e Stop introduzido na posição 192aa	BI520194
Exon 6	c.661delC	Thr188del fs e Stop introduzido na posição 200aa	BE789469
Exon 6	c.661delC + c.690delC	Thr188del fs e Stop introduzido na posição 201aa	BG430120
Exon 6	c.617delC	Phe174del e Stop introduzido na posição 175aa	BG433205
Exon 6	c.617delC	Phe174del e Stop introduzido na posição 175aa	BE870710
Exon 6	c.658delC		
	+677_683delTGGAC	Ala187del fs e Stop introduzido na posição 201aa	BQ224700
Exon 6	c.651delT	Thr185del fs e Stop introduzido na posição 200aa	BF130637
Exon 6	c.617delC	Phe174del e Stop introduzido na posição 175aa	BG432521
Exon 6	c.634delA		BE903556
Exon 6	c.722_724delCAG	Ser209del	BE785703

Exon 6	c.720delC	His208del fs	BI859911
Exon 6	c.713_717delTGGTG	Ser205del fs	BM908005
Exon 7	c.784delT +833delT	Leu229del fs	BI091624
Exon 7	c.763delC	Ala222del fs	BG284981
Exon 7	c.804delG	Asp236del fs	BX338304
Exon 7	c.812_816delAAATG	Arg238del fs	BQ877745
Exon 7	c.784delT +790delT	Leu229del fs	BE876057
Exon 7	c.784delT +790delT	Leu229del fs	BE874132
Exon 7	c.814delA	Asn239del fs	BX335972
Exon 7	c.780_781delAT + c.823delG	Ile228del fs	BX378078
Exon 7	c.833delT	Phe245del fs	BG179480 BF130637 BE788584 BE730063 AU124334 DN999234
Exon 8	c.926delA	Asn276del fs e Stop introduzido na posição 283aa	CD612875
Exon 8	c.882delU	Phe261del fs e Stop introduzido na posição 266aa	BE785436
Exon 8	c.882delU	Phe261del e Stop introduzido na posição 266aa	BE874132
Exon 8	c.901delA	Arg270del fs e Stop introduzido na posição 237aa	AL569316
Exon 8	c.931delC	Pro281del e Stop introduzido na posição 283aa	AL542222
Exon 8	c.931delC	Pro281del e Stop introduzido na posição 283aa	BX3999933
Exon 8	c.931delC	Pro281del e Stop introduzido na posição 283aa	BG542865
Exon 8	c.924_925delCA+ 931_932delCA	Asn276del fs e Stop introduzido na posição 283aa	BX357941
Exon 8	c.944delT	Gln286del fs	BX417634
Exon 8	c.953delT	Val290del fs	AL517809
Exon 8	c.925_931delAACAATT	Asn276del fs e Stop introduzido na posição 283aa	BE542347
Exon 8	c.924delC +931delC+938_939delAG	Asn276del fs	BX338737
Exon 8	c.906_908delCGA	Ala269del + Arg279Gly	H04403
Exon 8	c.924delC +931delC+943delC	Asn276del fs	AL551578

Tabela 7. Adicionais polimorfismo no gene ECE 2 Isoforma A e C.

Localização	Variações preditas de nucleotídeos em seqüências codificantes e não codificantes	Variações preditas de aminoácidos. fs: frameshift	ESTs
Exon 1/ ISO A	c.101_102delCG	Glu22del fs	AL520954
	c.140delC	Asp35del fs e Stop introduzido na posição 52aa	BG483855
Exon 2/ ISO A	c.[247delC +470_478del GACCGTGTC]	Leu 117 del fs	DA732003
	c.354_357delCAGC	Gln106del fs	AL520953
	c.452_457delTGGGGA	Gly Glu 139_140del e Stop introduzido na posição 141aa	BF035953
Exon 15/ ISO A	c.1845delC	Arg603del fs	BF511209
	c.1851delG	Ala605del fs	AW341249
	c.2019delCAGC	Gln661del fs	AA44831
Exon 22/ ISO A	[3073delT + 3099delG +3130delG+3140delC] 3'UTR		H46357
Exon 3/ ISO C	956_958delTGA 3'UTR		BG219927
	c.[804delC] + 837delC+851delC+869_870delTC	Ile253del fs	BI196548
	c.576delC	His180del +194 Stop Codon	BI199326
	c.576delC	His180del +194Stop Codon	BG336695
	c.[801delC] + 829_830delCT	Leu255del fs	BI761287
	c.795_780delTAAGT	Leu250del fs	BU188747
	c.738_743delCTCCCA	Pro Pro 234_235 del	CR997279
	c.[699delC+717delC+739delC+746delC+780delG+784_785delTC]	Gln221del fs	BE797471
	c.[699delC+723delC+746delC+754delT]	Gln221del fs	BF316745
	c.578delC	Phe181Phe 182Arg>Gly fs	DA732003
	c.578delC	Phe181Phe 182Arg>Gly fs	BG498979
	c.621delC	Stop codon 195aa	BE787675
	c.[621delC+623delG]	Stop codon 195aa	BG483855
	c.618delT	Ser194Pro e Stop introduzido na posição195aa	BU568980
	c.706delC	Ala223del	BE263969

Tabela 8. Adicionais polimorfismo no gene CXCR4 isoforma B.

Localização	Variações preditas de nucleotídeos em seqüências codificantes e não codificantes	Variações preditas de aminoácidos. fs: frameshift	ESTs
Exon 2b	c.[237_239delATC + 629_630delTG]	Ile48del + Ser>Arg178 fs	BI819797
	c.656_660delCCGCT	Arg188del fs +200 Stop codon	CD987769
	c.[802_803delAG + 854_857delGCCT]	Lys236del fs	BI754094
	c.[853_855delTGC + 871_873delTCA]	Leu253del + Ile259del	CD389217
	c.[9856delC +860_865delCTACAT +866T>C]	Leu253del fs	BQ718617
	c.885_889delTTCAT	Phe264del fs	BI762229
	c.901_902delTC	Ile269del fs + 274 Stop Codon	BX418530
	c.[989delC + 1016delC]	Pro>Pro299 fs +308 Stop Codon	BX421511
	c.975delC	His294del fs + 296 Stop Codon	BX353253
	c.1142delT	His350del fs	BI821693
	c.377_384delGTCTGGGAC	Ser30del fs +47 Stop Codon	CD251240

7. Referências Bibliográficas

ABBAS, K. A.; LICHTMAN, A. H.; POBER, J. S. *Imunologia celular e molecular*. 5ª ed. USA, **Saunders Company**, p. 60-61, 2005.

ADAMS, M. D.; KELLEY, J. M.; GOCAYNE, J. D.; DUBNICK, M.; POLYMERPOULOS, M. H. et al. **Complementary DNA sequencing: expressed sequence tags and human genome project**. *Science*, v.21, p. 1651-1656, 1991.

AKIYAMA, H.; BARGER, S.; BARNUM, S.; BRADT, B.; BAUER, J. et al. Inflammation and Alzheimer's disease. **Neurobiol. Aging**, v. 21, p. 383–421, 2000.

ALBRIGHT, A. V.; SHIEH, J. T.; ITOH, T.; LEE, B.; PLEASURE, D.; O'CONNOR, M. J.; DOMS, R. W.; GONZALEZ-SCARANO, F. Microglia express CCR5, CXCR4, and CCR3, but of these, CCR5 is the principal coreceptor for human immunodeficiency virus type 1 dementia isolates, **J. Virol**, v.73, p.205–213, 1999.

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment tool. **Journal of Molecular Biology**, v. 215, p. 403-410, 1990.

BARKER, G.; BATLEY, J.; O'SULLIVAN, H.; EDWARDS, K. J.; EDWARDS, D. Redundancy based detection of sequence polymorphisms in expressed sequence tag data using auto SNP. **Bioinformatics**, v.19, p.421-422, 2003.

BEREZIKOV, E.; CHUNG, W. J.; WILLIS, J.; CUPPEN, E.; LAI, E. C. .Mammalian mirtron genes. **Mol Cell**, v. 28, p.328–336, 2007.

BERTRAM, L.; MCQUEEN, M. B.; MULLIN, K.; BLACKER, D.; TANZI, R. E. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. **Nat Genet**, v. 39, p.17-23, 2007.

BLALOCK, E. M.; GEDDES, J. W.; CHEN, K. C.; PORTER, N. M.; MARKESBERY, W. R.; LANDFIELD, P.W. Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proceedings of the National Academy of Sciences*, v.101, p.2173-2178, 2004.

BOND, J.; SCOTT, S.; HAMPSHIRE, D. J.; SPRINGELL, K.; CORRY, P. et al. Protein-truncating mutations in ASPM cause variable reduction in brain size. **Am J Hum Genet**, v. 73 p. 1170-1177, 2003.

BORRONI, B.; COLCIAGHI, F.; PASTORINO, L.; ARCHETTI, S.; CORSINI, P. ApoE genotype influences the biological effect of donepezil on APP metabolism in Alzheimer disease: evidence from a peripheral model. **Eur Neuropsychopharmacol**, v.12, p.195-200, 2002.

BRAGA-NETO, U. M.; DOUGHERTY, E. R. Bolstered error estimation. **Pattern Recognition** v. 37, p.1267–1281. 2004.

BRAGA-NETO, U. M.; DOUGHERTY, E. R. Classification. In **Genomic Signal Processing and Statistics**. EURASIP Book series on Signal Processing and Communication, Hindawi Publishing Corporation. 2005.

BROOKS, W. M.; LYNCH P. J.; INGLE, C. C.; HATTON, A.; EMSON, P. C.; FAULL, R. L. M.; STARKEY, M. P. Gene expression profiles of metabolic enzyme transcripts in Alzheimer's disease. **Brain Research**, v. 1127, p. 127-135. 2007.

CHUZHANOVA, N.; COOPER, D. N.; FÉREC, C.; CHEN, J-N. Searching for potential microRNA-binding site mutations amongst know disease associated 3'UTR variants. **Genominc Med**. v. 1, p. 29-39, 2007.

CLOWER, D. M.; WEST, R. A.; LYNCH, J. C.; STRICK, P. L. *The Inferior Parietal Lobule Is the Target of Output from the Superior Colliculus, Hippocampus, and Cerebellum*. The Journal of Neuroscience, v.21, p. 6283-6291, 2001.

COLANGELO, V.; SCHURR, J.; BALL, M. J.; PELAEZ, R. P.; BAZAN, N. G.; LUKIW, W. J. Gene Expression Profiling of 12633 Genes in Alzheimer Hippocampal CA1: Transcription and Neurotrophic Factor Down-Regulation and Up-Regulation of Apoptotic and Pro-Inflammatory Signaling. **Journal of Neuroscience Research**, v.70 p. 462–473. 2002.

COPPOLA, G.; KARYDAS, A.; RADEMAKERS, R.; WANG, Q.; BAKER, M.; HUTTON, M.; MILLER, B. L.; GESCHWIND, D. H. **Annals of Neurology**, v.64, p. 92-96, 2008.

CORTINI. F. ; FENOGLIO, C. ; GRIDI I., VENTURELLI E., PONATI S. Novel exon 1 progranulin gene variant in Alzheimer's disease. **European Journal of Neurology**. v.15, p. 1111-1117, 2008.

DA SILVA, V. C.; RAMOS, F. J.; FREITAS, E. M.; BRITO MARQUES, P. B.; CAVALCANTI, N. M. H et al. Alzheimer's disease in Brazilian elderly has a relation with homocysteine but not with MTHFR polymorphisms. **Arquivos de Neuro-Psiquiatria**, v. 64, p. 941–945. 2006.

DEHMELT. L.; HALPAIN, S. The MAPS/Tau family of microtubule-associated proteins. **Genome Biology**, v 6, p.204-214, 2004.

DUBOIS, B.; FELDMAN, H. H.; JACOVA, C.; DEKOSKY, S. T.; BARBERGER-GATEAU, P et al., Research criteria for the diagnosis of Alzheimer's Disease: Revising the nincds–adrd criteria. **Lancet Neurol**, v. 6, p. 734–746, 2007.

ECKMAN, E. A; REED, D. K; ECKMAN, C. B. Degradation of the Alzheimer's amyloid beta peptide by endothelin-converting enzyme. **J. Biol. Chem**, v. 276, p. 24540–24548, 2001.

ECKMAN, E. A.; WATSON, M.; MARLOW, L.; SAMBAMURTI, K.; ECKMAN, C. B. Alzheimer's disease beta-amyloid peptide is increased in mice deficient in endothelin-converting enzyme. **J. Biol. Chem**, v. 278, p. 2081–2084, 2003.

EMILSSON, L.; SAETRE, P.; JAZIN, E. Alzheimer's disease: mRNA expression profiles of multiple patients show alterations of genes involved with calcium signaling. **Neurobiology of Disease**, v 21, p. 618 – 625, 2006.

FERNANDEZ, L. L.; SCHEIBE, R. M. Is MTHFR polymorphism a risk factor for Alzheimer's disease like APOE? **Arquivos de Neuro-Psiquiatria** v. 63, p. 1–6, 2005.

FORLENZA, O. V. Tratamento farmacológico da doença de Alzheimer. **Rev. psiquiatr. Clín**, v.32, p.137-148, 2005.

FORMENT, J.; GILABERT, F.; ROBLES, A.; CONEJERO, V.; NUEZ, F.; BLANCA, J. M. EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration. **BMC Bioinformatics**, v. 9, p. 1-10, 2008.

FREDMAN, D.; MUNNS, G.; RIOS, D.; SJOHOLM, F.; SIEGFRIED, M et al. HGVbase: a curated resource describing human DNA variation and phenotype relationships. **Nucleic Acids Res.** v.32, p.516- 519, 2004.

FREUDENBERG-HUA, Y.; FREUDENBERG, J.; KLUCK, N.; CICHON, S.; PROPPING, P.; NÖTHEN, M. M. Single nucleotide variation analysis in 65 candidate genes for CNS disorders in a representative sample of the European population. **Genome Res**, v.13 p.2271-2276, 2003.

GESCHWIND, D. H.; LOGINOV, M.; STERN, J. M. Identification of a locus on chromosome 14q for idiopathic basal ganglia calcification (Fahr disease). **Am J Hum Genet**, v. 65, p. 764-772, 1999.

GESCHWIND, D. H. Mice, Microarrays, and the genetic diversity of the brain. **Pnas**, v. 97 (20), P. 10676-10678, 2000.

GIBSON, G.; MUSE, S. V. SNPs and Variation. **A primer of Genome Science**. 2 ed, Saunderland, Massachusetts, p.133-137, 2004.

GRIFFITHS-JONES, S.; SAINI, H. K.; VAN DOGEN, S.; ENRIGHT, H. J. miRBASE: tools for microRNA genomics. **Nucleic Acids Research**, v. 36, p. 154-158, 2007.

GRIFFIN, W. S.; SHENG, J. G.; ROYSTON, M. C.; GENTLEMAN, S. M.; MCKENZIE, J. E et al. Glial-neuronal interactions in Alzheimer's disease: the potential role of a 'cytokine cycle' in disease progression. **Brain Pathol**, v. 8, p. 65–72, 1998.

GRIFFIN, W. S. T. IL-1 and the cytokine cycle in Alzheimer's disease. **J. Neurochem**, v.74, S52, 2000.

GUL, A.; HASSAN, M. J.; HUSSAIN, S.; RAZA, S. I.; CHISHTI, M. S.; AHMAD, W. A novel deletion mutation in CENPJ gene in a Pakistani family with autosomal recessive primary microcephaly. **J Hum Genet**, v. 51 p. 760-764, 2006.

HAASS, C.; SCHLOSSMACHER, M. G.; HUNG, A. Y.; VIGO-PELFREY, C.; MELLON, A et al. Amyloid beta-peptide is produced by cultured cells during normal metabolism. **Nature**, v 359, p. 322-325, 1992.

HACIA, J. G. Resequencing and mutational analysis using oligonucleotide microarrays. **Nature Genetics**, v. 21, p. 42-47. 1999.

HEESE, K.; AKATSU, H. Alzheimer's Disease - An Interactive Perspective. **Curr Alzheimer Res** v. 2, p.109-121, 2006.

HENEKA, M. T.; O' BANION M. K. Inflammatory processes in Alzheimer's disease. **Journal of neuroimmunology** v, 184, p.69-91, 2007.

HIRAKAWA, M.; TANAKA, T.; HASHIMOTO, Y.; KURODA, M.; TAKAGI, T.; NAKAMURA, Y. JSNP: a database of common gene variations in the Japanese population. **Nucleic Acids Res**, v. 30, p. 158-162, 2002.

HIRSCH-REINSHAGEN, V.; BURGESS, B. L.; WELLINGTON, C. L. Why lipids are important for Alzheimer disease? **Mol Cell Biochem**. In press DOI 10.1007/s11010-008-0012-2.(2008).

HU, J.; YAN, C. Identification of deleterious non-synonymous single nucleotide polymorphisms using sequence-derived information. **BMC Bioinformatics**, v. 9:297, 2008.

IMAIZUMIN, I. T.; MATSUMIYA, T.; YOSHIDA, H.; NARAOKA, T.; UESATO, R et al. Tumor-necrosis factor- α induces retinoic acid-inducible gene-I in rheumatoid fibroblast-like synoviocytes. **Immunology Letters**, (In Press), 2008.

IQBAL, K.; FLORY, M.; KHATOON, S.; SOININEN, H.; PIRTTILA, T et al.. Subgroups of Alzheimer's disease based on cerebrospinal fluid molecular markers. **Ann. Neurol**, v.58, p. 748–757, 2005.

KAUFMANN, W. A.; BARNAS, U.; HUMPEL, C.; NOWAKOWSKI, K.; DECOL, C et al. Synaptic loss reflected by secretoneurin-like immunoreactivity in the human hippocampus in Alzheimer's disease. **Eur J Neurosci**, v 10, p. 1084 –1094, 1998.

KAWASHIMA, S.; OGATA, H.; KANEHIS, A. M. AAindex: Amino Acid Index Database. **Nucl Acids Res**, v.27, p.368-369, 1999.

KIDD, J. M.; COOPER, G. M.; DONAHUE, W. F.; HAYDEN, H. S.; SAMPAS, N et al. Mapping and sequencing of structural variation from eight human genomes. Nature, v. 453, p. 56-64, 2008.

KIM, S.; DOUGHERTY, E. R.; SHMULEVICH, I.; HESS, K. R.; HAMILTON, S. R et al. Identification of Combination Gene Sets for Glioma Classification. **Molecular Cancer Therapeutics**, v 1, p.1229-1236, 2002.

KONISHI, Y.; BEACH, T.; SUE, L. I.; HAMPEL, H.; LINDHOLM, K.; SHEN, Y. The temporal localization of frame-shift ubiquitin-B and amyloid precursor protein, and

complement proteins in the brain of non-demented control patients with increasing Alzheimer's disease pathology. **Neuroscience Letters**, v. 348, p. 46–50, 2003.

LANDER, E. S.; LINTON, L. M.; BIRREN, B.; NUSBAUM, C.; ZODY, M. C, et al. Initial sequencing and analysis of the human genome. **Nature**, v. 15, p.860-921, 2001.

LEAL, G. F.; ROBERTS, E.; SILVA, E. O.; COSTA, S. M. R.; HAMPSHIRE, D. J.; WOODS, C. G. A novel locus for autosomal recessive primary microcephaly maps to 13q12.2. **J Med Genet**, v. 40, p. 540–542, 2003.

LEE, V. M.; GOEDERT, M.; TROJANOWSKI, J. Q. Neurodegenerative tauopathies. **Annu. Rev. Neurosci** v 24, p. 1121–1159, 2001.

LEMOS, R. R.; CASTELLETTI, C. H.; LIMA FILHO, J. L.; MARQUES, E.T.; OLIVEIRA, J. R. M. In silico identification of new genetic variations as potential risk factors for Alzheimer's Disease in a Microarray oriented simulation. **J. Mol. Neurosc**, In Press, DOI: 10.1007/s12031-009-9191-x.

LESK, A. M. Introdução à Bioinformática. 2 ed. Editora Artmed, Porto Alegre, 2008.

LUI, W-O.; POURMAND, N.; PATTERSON, B. K.; FIRE, A. Patterns of Known and Novel Small RNAs in Human Cervical Cancer. **Cancer Research**, v. 67, p. 6031-6043, 2007.

MATTSON, M. P. Ways towards and away from Alzheimer's disease. **Nature Res**, v.45, 38-78. 2004.

MILLER, J. A.; OLDHAM, M. C.; GESCHWIND, D. H. A systems level analysis of transcriptional changes in Alzheimer's Disease and normal aging. **The Journal Neuroscience**, v. 28, 1410-1420, 2008.

NAVRATIL, V.; PENEL, S.; DELMOTTE, S.; MOUCHIROUD, D.; GAUTIER, C.; AOUACHERIA, A. DigiPINS: A database for vertebrate exonic single nucleotide polymorphisms and its application to cancer association studies. **Biochimie**, v. 90, p.563-569, 2008.

NICKERSON, D. A.; TAYLOR, S. L.; WEISS, K. M.; CLARK, A. G.; HUTCHINSON, R. G et al. DNA sequence diversity in a 9.7kb region of the human lipoprotein lipase gene. **Nat. Genet**, v. 9, p. 233–240, 1998.

NICOLL, J. A. R.; MRAK, R. E.; GRAHAM, D. I.; STEWART, J.; WILCOCK, G et al. Association of interleukin-1 gene polymorphisms with Alzheimer's disease. **Ann. Neurol**, 47, 365–368, 2000.

NICHOLAS, A.; WOODS, G.; SWANDSON, E.; COX, J.; KARBANI, G. The molecular landscape of ASPM mutations in primary microcephaly. **J. Med. Genet**. In Press, DOI: 10.1136/jmg.2008.062380.

NISHIMURA, A. L.; GUINDALINI, C.; OLIVEIRA, J. R. M.; NITRINI, R.; BAHIA, V. S et al. Monamine Oxidase A polymorphism in brazilian patients : Risk factor for Late onset Alzheimer's disease?. **J Mol Neuroscience**, v. 7, p. 213-217, 2005.

NISHIMURA, A. L.; OLIVEIRA, J. R. M.; MATIOLI, S.; BRITO-MARQUES, P. R.; BAHIA, V. S.; NITRINI, R.; ZATZ, M. Analysis of the disease risk loci DXS1047 polymorphism in Alzheimer brazilian patients. **Molecular Psychiatry**, v. 5, p. 563-566. 2000.

NISHIMURA, A. L.; OLIVEIRA, J. R. M.; MITNE-NETO, M.; GUINDALINI, C.; NITRINI, R et al. Lack of Association Between the Brain-Derived Neurotrophin Factor (C-270T) Polymorphism and Late-Onset Alzheimer's Disease (LOAD) in Brazilian Patients. **J Mol Neuroscience**, v. 22, p. 257-260, 2004.

NISHIMURA, A. L.; OLIVEIRA, J. R. M.; OTTO, P. A.; MATIOLI, S. R.; BRITO-MARQUES, P. R et al. No evidence of association between the D10S1423 locus and Alzheimer disease in Brazilian patients. **J Neural Transm**, v. 108, p. 305-310, 2001.

OLIVEIRA, J. R. M.; GALLINDO, R. M.; MAIA, L. G. S.; BRITO-MARQUES, P. R.; OTTO, P. A et al. The short variant of the polymorphism within the promoter region of the serotonin transporter gene is a risk factor for late onset Alzheimer's Disease. **Molecular Psychiatry**, v. 3, p. 438-441, 1998.

OLIVEIRA, J. R. M.; NISHIMURA, A. L.; LEMOS, R. R.; ZATZ, M. The Genetics of Disease in Brazil: 10 years of Analysis in a Unique Population. **J. Mol. Neurosci**, v. 37, 74-79, 2009.

OLIVEIRA, J. R. M.; LEMOS, R. R. Linkage studies in familial idiopathic basal ganglia calcification: Separating the wheat from the chaff. **American Journal of Medical Genetics (Neuropsychiatric genetics)**. In Press. DOI 10.1002/ajmg.b.30839. 2008.

OLIVEIRA, J. R. M.; LIMA-FILHO, J. L.; SHIMOKOMAKI, C. M.; OKUMA, M.; PASSOS-BUENO, M. R.; ZATZ, M.; BRITO-MARQUES, P. R. The Use of Apolipoprotein E Genotype for Preclinical Detection of Risk's Group for Alzheimer's Disease. **American Journal of Medical Genetics (Neuropsychiatric Genetics)**, v. 74, p. 216-217, 1997.

OLIVEIRA, J. R. M.; SOBRIDO, M. J.; SPIFERI, E.; HOPFER, S.; MERONI, G et al. Analysis of Candidate Genes at the IBGC1 Locus Associated with Idiopathic Basal Ganglia Calcification ("Fahr's Disease"). **J. Mol. Neuroscience**, v. 33, p. 151-154, 2007.

OLIVEIRA, J. R. M.; SHIMOKOMAKI, C. M.; BRITO-MARQUES, P. R.; GALLINDO, R. M.; OKUMA, M et al. The association of the short polymorphism of the 5-HTTLPR polymorphism and the ApoE ε4 allele does not increase the risk for late onset Alzheimer's Disease. **Molecular Psychiatry**, v. 4, p. 19– 20, 1999.

PASTORINO, L.; SUN, A.; LU, P. J.; ZHOU, X. Z.; BALASTIK, M et al. The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-beta production. **Nature**, v. 440, p. 528-534, 2006.

PEREZ-IRATXETA, C.; BORK, P.; ANDRADE, M. Association of genes to genetically inherited diseases using data mining. **Nature genetics**, v. 31, p. 316-319, 2002.

PICOULT-NEWBERG, L.; IDEKER, T. E.; POHL, M. G.; TAYLOR, S. L.; DONALDSON, M. A et al. Mining SNPs from EST databases. **Genome Res**, v 9, p.167-74, 1999.

POIRIER, J. Apolipoprotein E4, cholinergic integrity and the pharmacogenetics of Alzheimer's disease. **J Psychiatry Neurosci**, v.24, p. 147-153, 1999.

POIRIER, J.; SEVIGNY, P. Apolipoprotein E4, cholinergic integrity and the pharmacogenetics of Alzheimer's disease. **J Neural Transm**, v.53, p.199-207. 1998.

RENCER, A. C. Methods of Multivariate Analysis. Wiley New York. 2002.

ROSSLER, M.; ZARSKI, R.; BOHL, J.; OHM, T. G. Stage-dependent and sector-specific neuronal loss in hippocampus during Alzheimer's disease. **Acta Neuropathol (Berl)**, v 103, p. 363–369, 2002.

RUSINIAK, M. E.; YU, M.; ROSS, D. T.; TOLHURST, E. C.; SLACK, J. L. Identification of B94 (TNFAIP2) as a Potential Retinoic Acid Target Gene in Acute Promyelocytic Leukemia1. **Cancer Research**, v. 60, p.1824–1829, 2000.

SARMA, V.; WOLF, F. W.; MARKS, R. M.; SHOWS, T. B.; DIXIT, V. M. Cloning of a novel tumor necrosis factor- α -inducible primary response gene that is differentially expressed in development and capillary tube-like formation *in vitro*. **J. Immunol**, v. 148, p. 3302–3312, 1992.

SMALL, G. W.; ERCOLI, L. M.; SILVERMAN, D. H.; HUANG, S. C.; KOMO, S et al. Cerebral metabolic and cognitive decline in persons at genetic risk for Alzheimer's disease. **Proc Natl Acad Sci U S A**, v. 97, p. 6037-42, 2000.

SMITH, T. F.; WATERMAN, M. S. Identification of common molecular subsequences. **Journal of Molecular Biology**, v. 147, p. 195-197, 1981.

STEFANSSON, H.; RUJESCU, D.; CICHON, S.; PIETILÄINEN, O. P.; INGASON, A et al. Large recurrent microdeletions associated with schizophrenia. **Nature**. 2008 Jul 30. [Epub ahead of print].

STEINMAN, L. Nuanced roles of cytokines in three major human brain disorders. **The Journal of Clinical Investigation**, v.118, p.3557-3563, 2008.

ST GEORGE-HYSLOP, H. P. Piecing together Alzheimer's. **Sci Am**, v. 283, p. 76-83, 2000 (review).

TANG, J.; VOSMAN, B.; VOORRIPS, R. E.; GERARD VAN DER LINDEN, G.; LEUNISSEN, J. A. M. QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. **BMC Bioinformatics**, v. 7, p. 438-452, 2006.

TANG, J.; LEUNISSEN, J. A. M.; VOORRIPS, R.E.; VAN DER LINDEN, C.G.; VOSMAN, B. HaploSNPer: a web-based allele and SNP detection tool. **BMC Genetics**, v.9, p.1-7, 2008.

TATUSOVA, A.; MADDEN, T. L. Blast 2 sequences, a new tool for comparing protein and nucleotide sequences. **Microbiology Letters**, v. 174, p.247-250, 1999.

THORISSON, G. A.; STEIN, L. D. The SNP Consortium website: past, present and future, **Nucleic Acids Res**, v. 31, p. 124-127, 2003.

USECHE, F. J.; GAO, G.; HANAFEY, M.; RAFALSKI, A. High-Throughput Identification, Database Storage and Analysis of SNPs in EST Sequences. **Genome Informatics**, v. 12, p. 194–203, 2001.

VAN DE VIJVER, M. J.; HE, Y. D.; VAN'T VEER, L. J.; DAI, H.; HART, A. A. M et al. Amyloid metabolism and secretases in Alzheimer's disease. **Curr Neurol Neurosci Rep**, v. 1, p. 422-427, 2001.

VAN LEEUWEN, F. W.; DE KLEIJN, D. P. V.; VAN DEN HURK, H. H.; NEUBAUER, A.; SONNEMANS, M. A. F et al. Frameshift mutants of b amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. **Science**, v. 279, p. 242–247, 1998.

VEPSÄLLÄINEN, S.; HELISALMI, S.; KOIVISTO, A. M.; HILTUNEM, M.; SOINNEM, H. Somatostatin genetic variants modify the risk for Alzheimer's disease among Finnish patients. **J Neurol**, v.254, p.1504–1508, 2007.

XIA, M. Q.; HYMAN, B.T. Chemokines/chemokine receptors in the central nervous system and Alzheimer's disease. **J. NeuroVirol**, v.5, p. 32–41, 1999.

XIA, W. Amyloid metabolism and secretases in Alzheimer's disease. **Curr Neurol Neurosci Rep**, v. 1, p.44-427, 2001 (review)

WALSH, T.; MCCLELLAN, J. M.; MCCARTHY, S. E.; ADDINGTON, A. M.; PIERCE, S. B et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. **Science**, v 320, p 539-543, 2008.

WANG, D. G.; FAN, J. B.; SIAO, C.; BERNO, A.; YOUNG, P et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. **Science**, v. 280, p. 1077–1082, 1998.

WEERARATNA, A. T.; KALEHUA, A.; DELEON, I.; BERTAK, D.; MAHER, G et al. Alterations in immunological and neurological gene expression patterns in Alzheimer's disease tissues. **Experimental cell research**, v. 313, p. 450-461.2007.

WEISS, K. M.; TERWILLIGER, T. D. How many diseases does it take to map a gene with SNPs? **Nature Genetics**, v. 26, p.151-157, 2000.

WHEELER, D. L.; BARRETT, T.; BENSON, D. A.; BRYANT, S. H. ; CANESE, K et al. Database resources of the National Center for Biotechnology Information, **Nucleic Acids Res**, v. 33, p 39-45, 2005.

WINKLHOFER, K. F.; TATZELT, J.; HAASS, C. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. **EMBO J**, v. 27, p. 336-349.2008.

World Health Organization. <http://www.who.int>

WYSS-CORAY, T. Inflammation in Alzheimer disease: driving force, bystander or beneficial response? **Nature Medicine**, v. 129 p. 1005-1016, 2006.

YU, C. E.; DAWSON, G.; MUNSON, J.; D'SOUZA, I.; OSTERLING, J et al. Presence of large deletions in kindreds with autism. **Am J Hum Genet**. v. 71, p.100-115, 2002.

YU, W.; CLYNE, M.; DOLAN, S. M.; YESUPRIYA, A.; WULF, A et al. GAPscreener: an automatic tool for screening human genetic association literature in *PubMed* using the support vector machine technique. **BMC Bioinformatics**, v. 9, p. :205 2008.

YU, WEI.; WULF, A.; LIU, T.; KHOURY, M. J.; GWINN, M. Gene Prospector: An evidence gateway for avaluating potential susceptibility genes and interacting risk factores for human diseases. **BMC Bioibformatics**. v. 9: 528, 2008.