

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE CIÊNCIAS BIOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

Análise genética de novos potenciais polimorfismos de risco em Transtornos do  
Humor e utilização de abordagens computacionais em busca de genes candidatos a  
Doença de Alzheimer

Manuela Barbosa Rodrigues de Souza  
Orientador: Prof. Dr. João Ricardo Mendes de Oliveira

Recife  
2013

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE CIÊNCIAS BIOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

Análise genética de novos potenciais polimorfismos de risco em Transtornos do Humor e utilização de abordagens computacionais em busca de genes candidatos a Doença de Alzheimer

Tese apresentada ao programa de Pós-graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como parte dos requisitos para obtenção do grau de Doutor em Ciências Biológicas.

Manuela Barbosa Rodrigues de Souza  
Orientador: Prof. Dr. João Ricardo Mendes de Oliveira

Recife  
2013

Catálogo na Fonte:  
Bibliotecário Bruno Márcio Gouveia, CRB-4/1788

S729a Souza, Manuela Barbosa Rodrigues de

Análise genética de novos potenciais polimorfismos de risco em transtornos do humor e utilização de abordagens computacionais em busca de genes candidatos a Doença de Alzheimer / Manuela Barbosa Rodrigues de Souza. – Recife: O Autor, 2013.

97 f. : il., fig., tab.

Orientadora: João Ricardo Mendes de Oliveira

Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Ciências Biológicas. Pós-graduação em Ciências Biológicas, 2013.

Inclui bibliografia e anexos

1. Sistema nervoso – Doenças 2. Alzheimer, Doença de 3. Polimorfismos (genética) I. Oliveira, João Ricardo Mendes de (orientador) II. Título.

616.8

CDD (22.ed.)

UFPE/CCB-2013-122

**Análise genética de novos potenciais polimorfismos de risco em Transtorno do Humor e utilização de abordagens computacionais em busca de genes candidatos a Doença de Alzheimer**

Tese apresentada ao programa de pós-graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como parte integrante dos requisitos para obtenção do grau de Doutor em Ciências Biológicas.

**Data da Defesa:** 19 de abril de 2013.

**Banca Examinadora**

---

Prof. Dr. João Ricardo Mendes de Oliveira  
Orientador – Universidade Federal de Pernambuco

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Danyelly Bruneska Gondim Martins  
Universidade Federal de Pernambuco

---

Prof. Dr. Paulo Roberto Eleutério de Souza  
Universidade Federal Rural de Pernambuco

---

Dr. Carlos Henrique Madeiros Castelletti  
Instituto Agrônômico de Pernambuco

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Thaís Gaudencio do Rêgo  
Universidade Federal da Paraíba



“Por vezes sentimos que aquilo que fazemos  
não é senão uma gota de água no mar. Mas o  
mar seria menor se lhe faltasse uma gota.”

Teresa de Calcutá

## Agradecimentos

A Deus por iluminar meu caminho e me dar forças para seguir sempre em frente.

Ao Professor Dr. João Ricardo, especial agradecimento pela orientação do meu doutorado, pela paciência, amizade, tempo e atenção dedicados. Por ter me dado à liberdade de fazer parcerias em outros centros, o que foi fundamental para o meu crescimento.

Ao Professor Dr. Ivan Gesteira, por suas importantes contribuições e ensinamentos que foram fundamentais para o desenvolvimento desta tese. Seu profissionalismo, competência e seriedade foram essenciais para o meu aperfeiçoamento acadêmico, profissional e pessoal.

Ao Professor Dr. Paulo Roberto Eleutério de Souza e ao pesquisador Dr. Carlos Henrique Madeiros Castelletti pelas contribuições no exame de qualificação.

À Professora Dra. Nara Suzy Aguiar de Freitas, pelo incentivo, pela amizade e pelas revisões referentes ao trabalho.

Aos demais Professores, Pesquisadores e Colegas de turma, com os quais eu tive a oportunidade de cursar disciplinas, desenvolver pesquisas, em especial ao Prof. Dr. Sandro Gonçalves de Lima.

A Adenilda Eugênia, secretária da Pós-graduação em Ciências Biológicas, por todo apoio e por ser sempre tão solícita.

Ao amigo e companheiro de pesquisa Gilderlanio Santana, por toda divisão de dificuldades, crescimento e aprendizado na área de bioinformática.

Aos amigos do Grupo de Pesquisa Psiquiatria Biológica: Danyllo Oliveira, Isis Lima, Joana Braga, Jose Eriton, Karinna Teixeira, Lylyan Fragoso Pimentel, Regina Galdino, Roberta Lemos, e Samy Scherb Steinberg. Por todos momentos vividos juntos durante a pós-graduação.

Aos amigos do Centro de Informática: Fausto Lorenzato, Luciano Pacífico, por toda paciência, amizade e parceria na disciplina de Algoritmos e Métodos para Bioinformática e Biologia Computacional. E a Adenilton Silva, pela amizade, conselhos e incentivos.

Ao meu namorado Valmir Macário, que me ajudou de tantas formas que seria difícil relatar todas.

Aos meus familiares, pelo apoio e motivação.

A minha amiga Amanda Montalvão, pela correção da formatação da tese.

Aos membros da banca examinadora, por suas sugestões de melhoria para o enriquecimento do nosso trabalho.

À FACEPE – Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco -, pelo apoio financeiro durante o desenvolvimento desta tese.

A todos que, direta ou indiretamente, contribuíram para a conclusão desta importante etapa, meus sinceros agradecimentos.

## Resumo

Doenças neuropsiquiátricas afetam cerca de 450 milhões de pessoas em todo mundo e dentre estas patologias os Transtornos do Humor (TH) e Doença de Alzheimer (DA) são as mais comuns. Em relação à sua etiologia as doenças neuropsiquiátricas são resultados de variações em um grupo de genes e de fatores ambientais. Pesquisas recentes vêm mostrando associação positiva entre variações genéticas em genes envolvidos nos sistemas de neurotransmissores com o desenvolvimento de doenças neuropsiquiátricas. Por isso, os estudos sobre polimorfismos genéticos, nas doenças psiquiátricas são de grande importância para a compreensão dos mecanismos moleculares envolvidos e podem auxiliar no diagnóstico das mesmas. Neste cenário, mutações encontradas no DNA têm sido amplamente estudadas, a fim de elucidar aspectos genéticos relacionados às neuropatologias. Os polimorfismos do tipo *SNPs* (Polimorfismo de Base Única) e *INDELs* (inserções e deleções de fragmentos de DNA) têm se destacado devido as fortes associações com os TH e DA. Diversos métodos de biologia molecular têm sido utilizados para detectar estes tipos de polimorfismos, os experimentos moleculares geram grande quantidade de dados a serem analisados, fazendo-se necessário a utilização de ferramentas computacionais para se extrair informações a partir desses dados gerados. Assim, o objetivo desse estudo foi o uso de bioinformática e de genotipagem em larga escala na busca de novos polimorfismos genético em TH e aplicação de ferramentas computacionais em banco de dados de GWAS referente à DA. Para obter os resultados referentes à TH optamos por utilizar o software CLCbio Workbench<sup>®</sup>, sequenciamento automatizado mega Bace 1000 e experimentos preliminares da técnica de DNA *pooling*. Já para a DA, utilizamos o teste de associação e método de regressão linear do software PLINK e o pacote *genetics* da linguagem de programação R, para correlacionar os níveis da proteína  $\beta$ amiloide no plasma e líquido cefalorraquidiano e um total de 598.821 *SNPs*, ambos os dados oriundos do banco de dados ADNI (*Alzheimer's Disease Neuroimaging Initiative*). Após uma sequência de passos *in silico* identificamos variações anteriormente descritas e novos polimorfismos candidatos à fisiopatologia dos TH, na fase de validação dessas variações, por meio de sequenciamento, falsos positivos foram frequentemente identificados, sendo descartados após a verificação na cadeia complementar. Apenas o SNP rs14068, localizado no exon 2 do gene *GABRA5* foi validado em amostras de pacientes com TH. No estudo referente à DA, 5 *SNPs* nas regiões dos genes *TOMM40*, *PAMR1*, *TRIM9* e *CCDC112* e 3 *SNPs* em regiões de intron atingiram associação significativa, levantando a possibilidade de estejam relacionados a fisiopatologia da DA.

Palavras chaves: doenças neuropsiquiátricas, *SNPs*, *INDELs*, bioinformática, GWAS.

## Abstract

Neuropsychiatric diseases affect about 450 million people worldwide, among these pathologies the Mood Disorders and Alzheimer disease are the most common. Regarding its etiology neuropsychiatric disorders are the result of variations in a group of genes and environmental factors. Recent studies have shown a positive association between genetic variations in genes involved in neurotransmitter systems with the development of neuropsychiatric disorders. Therefore, studies of genetic polymorphisms in psychiatric diseases are of great importance for understanding the molecular mechanisms involved and may assist in the diagnosis of these diseases. In this scenario, mutations found in DNA have been widely studied in order to elucidate the genetic aspects related neuropathologies. Polymorphisms of type *SNPs* (*Single Nucleotide Polymorphism*) and *INDELs* (insertions and deletions) have been highlighted due to strong associations with Mood Disorders and Alzheimer's Disease. Several molecular biology methods have been used to detect these types of polymorphisms, however, the molecular experiments generate large amounts of data to be analyzed and which are deposited, often in public databases, making it necessary to use tools computer to extract information from these data. The objective of this study is the use of bioinformatics and large-scale genotyping in the search for new genetic polymorphisms in Mood Disorders and application of computational tools for database GWAS on the Alzheimer's Disease. For best results we chose to use the software CLCbio Combined Workbench<sup>®</sup>, sequencing Mega Bace 1000 and preliminary experiments the technique of DNA pooling. As for the DA used the association test and linear regression software package plink and genetics of the R programming language, to correlate protein levels  $\beta$ amiloide in plasma and cerebrospinal fluid and a total of 598,821 *SNPs*, both data from the ADNI database (Alzheimer's Disease Neuroimaging Initiative). After a sequence of steps in silico identified previously described variations and new candidate polymorphisms pathophysiology of TH, during validation of these variations, by sequencing, often false positives were identified and discarded after checking the complementary strand. Only the SNP rs14068, located in exon 2 of the gene GABRA5 was validated on samples of patients with Mood Disorders. In the study Alzheimer's disease, 5 *SNPs* in regions of genes *TOMM40*, *PAMR1*, *TRIM9* and *CCDC112* and 3 *SNPs* in intron regions reached significant association, raising the possibility that they are related to the pathophysiology of Alzheimer's disease.

Keywords: neuropsychiatric diseases, *SNPs*, *INDELs*, bioinformatics, GWAS.

| Lista de Figuras da Revisão Bibliográfica                                                                                                                                                                                                                                                                                                                                                                                                   | Página |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| Figura 1. Envolvimento individual e em conjunto dos neurotransmissores noradrenalina, serotonina e dopamina em algumas funções cognitivas.                                                                                                                                                                                                                                                                                                  | 17     |
| Figura 2. Representação geral dos locos cromossômicos dos genes envolvidos nos sistema de neurotransmissores glutamatérgico e gabaérgico, que apresentaram níveis de expressão alterada, de acordo com o estudo de <i>Microarray</i> de expressão, realizado por Choudary <i>et.al.</i> , 2005. Esses genes foram inseridos no nosso estudo de busca por novos potenciais polimorfismos genéticos de risco em TH.                           | 19     |
| Figura 3. Receptores ionotrópico de glutamato: NMDA, AMPA e Kainato. Distribuídos na membrana neuronal.                                                                                                                                                                                                                                                                                                                                     | 20     |
| Figura 4. Esquema de um receptor de GABA A formando o pentâmero $\alpha 2 \beta 2 \gamma$ .                                                                                                                                                                                                                                                                                                                                                 | 21     |
| Figura 5. Progressão da DA no cérebro. Placas e emaranhados, representados nas áreas azuis, tendem a se espalhar através do córtex em um padrão previsível, à medida que a DA progride, em uma taxa de progressão é variável entre os pacientes. Adaptado de: <a href="http://www.alz.org/braintour/progression.asp">http://www.alz.org/braintour/progression.asp</a> .                                                                     | 24     |
| Figura 6. GWAS publicados a partir de 2005 até junho de 2012, posicionados de acordo com a localização do locus estudado nos cromossomos humanos; os círculos coloridos representam as categorias das doenças estudadas, os círculos amarelo-claros englobam as doenças neuropsiquiátricas, incluindo a DA. Disponível em: <a href="http://www.genome.gov/GWAStudies">http://www.genome.gov/GWAStudies</a> (HINDORFF <i>et al.</i> , 2012). | 28     |
| Figura 7. Arquitetura de uma Rede Neural Artificial com os três grupos de camadas.                                                                                                                                                                                                                                                                                                                                                          | 31     |
| Figura 8. Página inicial de site do banco de dados ADNI.                                                                                                                                                                                                                                                                                                                                                                                    | 32     |
| Figura 9. Página inicial do site do software PLINK.                                                                                                                                                                                                                                                                                                                                                                                         | 34     |
| Figura 10. Página inicial do site do software GeneMANIA.                                                                                                                                                                                                                                                                                                                                                                                    | 36     |

|                                                                                                                                                                                                                                                             |         |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Lista de tabelas dos artigos                                                                                                                                                                                                                                | Página  |
| Capítulo 2 - Artigo 1                                                                                                                                                                                                                                       |         |
| Searching for New Genetic Variations in Expression Databases for the GABAergic and Glutamatergic Systems                                                                                                                                                    |         |
| Table 1. Summary of <i>in silico</i> analysis of <i>GABRA5</i> , <i>GRM3</i> , <i>GRIK5</i> , <i>GABRB3</i> , <i>GABRD</i> , <i>GLUL</i> , and <i>GABRG2</i> genes.                                                                                         | 51      |
| Table 2. Additional polymorphisms in <i>GABRA5</i> , <i>GRM3</i> , <i>GRIK5</i> , <i>GABRB3</i> , <i>GABRD</i> , <i>GLUL</i> , and <i>GABRG2</i> genes.                                                                                                     | 52 - 55 |
| Capítulo 3 - Artigo 2                                                                                                                                                                                                                                       |         |
| Exploring the Implications of <i>INDELs</i> in Neuropsychiatric Genetics: Challenges and Perspectives                                                                                                                                                       |         |
| Table 1. Divergence in both the number and size of <i>INDELs</i> reported in personal genome.                                                                                                                                                               | 62      |
| Table 2. <i>INDELs</i> in relevant genes associated neuropsychiatric diseases.                                                                                                                                                                              | 63      |
| Capítulo 4 - Artigo 3                                                                                                                                                                                                                                       |         |
| Genome-wide association analysis combined with CSF and Plasma Biomarkers (AB-40 & AB-42) highlight new risk factors for Alzheimer's Disease                                                                                                                 |         |
| Table 1. Details of experimental designs performed.                                                                                                                                                                                                         | 88      |
| Abbreviations: CSF, cerebrospinal fluid; Chr, chromosome; M12, one year late; AD, Alzheimer Disease; MCI, Mild Cognitive Impairment; SNP, single nucleotide polymorphism; MAF, minor allele frequency; SD, standard deviation; L, level; A, age; N, number. |         |
| Table 2. Top <i>SNPs</i> with $p\text{-value} < 10^{-4}$ . <i>SNPs</i> associated with CSF and Plasma biomarkers in diagnostic groups (NC, MCI and AD).                                                                                                     | 89      |
| Abbreviations: CSF, Cerebrospinal fluid; Chr, chromosome; M12, one year late; AD, Alzheimer Disease; MCI, Mild Cognitive Impairment; NC, normal control; SNP, single nucleotide polymorphism.                                                               |         |
| Table 3. Genotypic frequency.                                                                                                                                                                                                                               | 89      |
| Abbreviations: BL, baseline; M12, one year late; AD, Alzheimer Disease; MCI, Mild Cognitive Impairment; NC, normal control; N, number.                                                                                                                      |         |

| Lista de Figuras dos Artigos                                                                                                                                                                                                                                                                                                                  | Página |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| Capítulo 2 - Artigo 1                                                                                                                                                                                                                                                                                                                         |        |
| Searching for New Genetic Variations in Expression Databases for the GABAergic and Glutamatergic Systems                                                                                                                                                                                                                                      |        |
| Figure 1. Sequential steps of the proposed bioinformatics pipeline.                                                                                                                                                                                                                                                                           | 50     |
| Figure 2. Clusters of ESTs at the positions of some predicted variations for the genes GABRA5, GABRA3, GLUL, and GABRG2. Identification of ESTs codes and positions in UTR and CDS regions are fully displayed at Table 2.                                                                                                                    | 51     |
| Capítulo 3 - Artigo 2                                                                                                                                                                                                                                                                                                                         |        |
| Exploring the Implications of <i>INDELs</i> in Neuropsychiatric Genetics: Challenges and Perspectives                                                                                                                                                                                                                                         |        |
| Figure 1. A pipeline suggested to merge different data resources in order to combine expression studies and genetic variations and predicted to detect potential new polymorphisms, including <i>INDELs</i> .                                                                                                                                 | 64     |
| Capítulo 4 - Artigo 3                                                                                                                                                                                                                                                                                                                         |        |
| Genome-wide association analysis combined with CSF and Plasma Biomarkers (A $\beta$ -40 & A $\beta$ -42) highlight new risk factors for Alzheimer's Disease                                                                                                                                                                                   |        |
| Figure 1. Manhattan and Q-Q plots of the quantitative trait association analysis in AD and MCI diagnostic groups. A, A $\beta$ -42 CSF baseline AD diagnosis; B, A $\beta$ -42 CSF baseline MCI diagnosis; C, A $\beta$ -42 Plasma Baseline AD diagnosis; D, A $\beta$ -42 Plasma M12 AD diagnosis; E, A $\beta$ -40 Plasma M12 AD diagnosis. | 87     |
| Figure 2. Mean of biomarker levels as a function of CSF and plasma baseline and plasma M12 genotypes over all diagnostic groups (NC, MCI and AD) for the most significant associated <i>SNPs</i> .                                                                                                                                            | 88     |

## Lista de Abreviações

AMPA – Receptor Alfa-Amino-3-Hidroxi-5-Metil-4-Isoxazol Propiônico, do Inglês:  *$\alpha$ -Amino-3-Hydroxy-5-Methyl-4-Isoxazolepropionic Acid Receptor*

APOE – Apolipoproteína E, do Inglês: *Apolipoprotein E*

APP - Proteína Percussora Amiloide, do Inglês: *Amyloid Precursor Protein*

BLAST – Ferramenta Básica de Busca de Alinhamentos Locais, do Inglês: *Basic Local Alignment Search Tool*

CSF - Flúido Cerebroespinal, do Inglês: *Cerebrospinal Fluid*

DA - Doença de Alzheimer

DCJ - Doença de Creutzfeldt–Jakob

DCL - Demência por Corpos de Lewy

DFT - Demência Frontotemporal

DH - Demências Hidrocefálicas

DT - Demências Toxicometabólicas

DV - Demência Vascular

GABA - Ácido Gama-Aminobutírico, do Inglês: *Gamma-Aminobutyric Acid*

GWAS - Estudos de Associação que Englobam todo o Genoma, do Inglês: *Genome-wide Association Study*

INDELs - Inserções e Deleções, do Inglês: *Insertion and Deletion*

MCI - Comprometimento Cognitivo leve, do Inglês: *Mild Cognitive Impairment*

MRI - Ressonância Magnética, do Inglês: *Magnetic Resonance Imaging*

NMDA – Receptor N-metil-D-Aspartato, do Inglês: *N-Methyl-D-aspartat acid receptor*

PCR reação em cadeia da polimerase, do Inglês: *Polymerase Chain Reaction*

PET - Tomografia por Emissão de Pósitrons, do Inglês: *Positron Emission Tomography*

RNAs - Redes Neurais Artificiais

SAP - Polimorfismo de Aminoácido Único, do Inglês: *Single Aminoacid Polymorphism*

SNP - Polimorfismo de Base Única, do Inglês: *Single Nucleotide Polymorphism*

TB - Transtornos Bipolares

TDT - Teste de Desequilíbrio de Ligação, do Inglês: *Transmission Disequilibrium Test*

TH - Transtornos do Humor

TU - Transtorno Depressivo Unipolar

TYH - Tirosina Hidroxilase, do Inglês: *Tyrosine Hydroxylase*

VNTR - Repetição em Tandem em Número Variável, do Inglês: *Variable Number of Tandem Repeats*

$\beta$ A - Proteína Beta-Amiloide, do Inglês: *Beta-Amyloid Protein*



## Sumário

|                                                                                                                                                   |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1. Introdução.....                                                                                                                                | 13 |
| Capítulo 1                                                                                                                                        |    |
| 2. Revisão Bibliográfica                                                                                                                          |    |
| 2.1. Transtornos do Humor.....                                                                                                                    | 16 |
| 2.2. Fatores de risco genético associado aos TH.....                                                                                              | 17 |
| 2.3. Genes dos sistemas glutamatérgico e gabaérgico inseridos no estudo relacionado aos TH.....                                                   | 18 |
| 2.3.1. Gene <i>GLUL</i> .....                                                                                                                     | 19 |
| 2.3.2. Genes <i>GRIK5</i> e <i>GRM3</i> .....                                                                                                     | 19 |
| 2.3.3. Genes <i>GABRB3</i> , <i>GABRD</i> , <i>GABRA5</i> e <i>GABRG2</i> .....                                                                   | 20 |
| 2.4. Demências e Doença de Alzheimer.....                                                                                                         | 22 |
| 2.5. Genética na Doença de Alzheimer.....                                                                                                         | 25 |
| 2.6. Polimorfismos Genéticos.....                                                                                                                 | 26 |
| 2.7. Estudos Genéticos de Associação.....                                                                                                         | 27 |
| 2.8. Bioinformática.....                                                                                                                          | 29 |
| 2.8.1. Alinhamento de Sequências.....                                                                                                             | 29 |
| 2.8.2. Redes Neurais Artificiais e Software PMut.....                                                                                             | 30 |
| 2.8.3. Banco de dados Alzheimer's Disease Neuroimaging Initiative – ADNI .....                                                                    | 32 |
| 2.8.4. PLINK .....                                                                                                                                | 33 |
| 2.8.5. GeneMANIA .....                                                                                                                            | 35 |
| 3. Referências Bibliográficas.....                                                                                                                | 37 |
| Capítulo 2 – Artigo 1 .....                                                                                                                       | 48 |
| Searching for New Genetic Variations in Expression Databases for the GABAergic and Glutamatergic Systems.....                                     | 48 |
| Apêndice do Capítulo 2 .....                                                                                                                      | 58 |
| Troubleshooting the search for potential pathogenic small <i>INDELs</i> (1-10bp) in neuropsychiatric disorders.....                               | 58 |
| Capítulo 3 – Artigo 2.....                                                                                                                        | 60 |
| Exploring the Implications of <i>INDELs</i> in Neuropsychiatric Genetics: Challenges and Perspectives .....                                       | 60 |
| Capítulo 4 – Artigo 3.....                                                                                                                        | 67 |
| Genome-wide association analysis combined with CSF and Plasma Biomarkers (AB-40 & AB-42) highlight new risk factors for Alzheimer's Disease ..... | 67 |
| Conclusão .....                                                                                                                                   | 93 |
| Anexos.....                                                                                                                                       | 96 |

## 1. Introdução

As doenças neuropsiquiátricas afetam cerca de 450 milhões de pessoas em todo mundo e são responsáveis por 14% da carga global de doenças, 30% de todas as doenças não transmissíveis e representam mais de 10% de incapacidade em todo o mundo. Os transtornos do humor (TH) são as principais causa de incapacidade em jovens e a doença de Alzheimer (DA) é a responsável pelas maiores causas de incapacidade em idosos. Estimativas mostram que a prevalência das doenças neuropsiquiátricas deverá continuar a crescer nos próximos anos. Principalmente nos grandes centros urbanos, onde a população está sujeita a diversos fatores estressores. Estima-se que em 2030, a depressão seja a segunda causa na carga global de doença, e em 2040 aproximadamente 81,1 milhões de pessoas viverão com alguma forma de demência.

Dentre os polimorfismos genéticos que influenciam na susceptibilidade a doenças, os polimorfismos de base única (*SNPs*, do Inglês: *Single Nucleotide Polymorphisms*) e as inserções e deleções (*INDELs*), por apresentarem potencial impacto tanto nas atividades proteicas quanto nas alterações de níveis de expressão gênica, têm sido cada vez mais reconhecidos como os principais tipos de variações genéticas em traços genéticos complexos, como no caso dos TH e da DA. Além de serem os mais frequentes polimorfismos encontrados ao longo do DNA, ocorrem uma vez em cada 300 nucleótidos, em média, o que significa que existem cerca de 10 milhões de *SNPs* no genoma humano, representando a classe mais frequente de variações genéticas, seguida dos *INDELs*.

Apesar do envolvimento de polimorfismos genéticos ser fortemente sugerido, as relações entre os polimorfismos e os aspectos funcionais nos TH e na DA ainda não estão compreendidas, consequentemente, um dos grandes desafios da genética molecular humana é identificar e compreender como as mudanças no DNA afetam o risco dos indivíduos desenvolverem estas doenças e criar métodos mais efetivos de prevenção, diagnóstico e tratamento. Portanto, a realização de estudos sobre polimorfismos genéticos nas doenças neuropsiquiátricas, tais como TH e DA, é de importância estratégica, haja vista o aumento progressivo e contínuo destes transtornos em todo o mundo. Tais estudos são fundamentais para que se possam identificar potenciais novos fatores genéticos de risco que auxilie na compreensão dos mecanismos moleculares envolvidos na fisiopatologia dessas doenças, podendo assim auxiliar no diagnóstico e no planejamento de terapias dessas patologias.

Por agregar conhecimentos na área da estatística, informática e biologia, a bioinformática tem sido bastante utilizada para dar suporte às análises moleculares e tem sido de grande auxílio na busca de polimorfismos genéticos que possam estar associado a doenças complexas, como TH e DA. Diante deste cenário, nosso trabalho teve como objetivo principal o estudo de polimorfismo

genético em TH através do uso de bioinformática e de genotipagem em larga escala, priorizando genes ligados aos sistemas de neurotransmissor gabaérgico e glutamatérgico, e aplicação de ferramentas computacionais em banco de dados de GWAS (do Inglês: *Genome-wide Association Study*) referente à DA. Como objetivos específicos este trabalho foram realizadas triagem e validações das variações previstas no estudo *in silico* relacionado com TH, em amostras de DNA de pacientes da população brasileira; e utilizadas ferramentas computacionais na produção de um painel de polimorfismos associados aos diferentes níveis do biomarcador  $\beta$  amiloide, no plasma sanguíneo e no líquido cefalorraquidiano, em grupos de pacientes com transtorno cognitivo leve e DA.

Assim, este trabalho está organizado da seguinte forma:

No capítulo 1 apresentamos uma revisão do estado da arte onde são introduzidos conceitos básicos para o entendimento do nosso trabalho. São fornecidas informações relevantes sobre os TH e DA, são tratados os conceitos de *SNP* e *INDELs*, e são apresentados o banco de dados e os programas computacionais utilizados no nosso trabalho.

O capítulo 2 trata da busca de novas variações genéticas nos sistemas gabaérgico e glutamatérgico de neurotransmissores por meio de método *in silico* e da possibilidade de que deleções de até 7 pares de bases, em genes dos referidos sistemas de neurotransmissor, estejam ligadas à fisiopatologia dos TH. Complementando esse capítulo, mostramos resultados preliminares da validação, por método de sequenciamento direto na população brasileira, dos polimorfismos encontrados *in silico*. Também citamos a realização de experimentos preliminares por meio da técnica de DNA *pooling*, ressaltando a necessidade de ajustes metodológicos para melhor interpretação do resultado, mas este método é potencialmente eficaz para economizar tempo e recursos na triagem de polimorfismos em larga escala.

No capítulo 3 exploramos os impactos dos *INDELs* nas doenças neuropsiquiátricas, os desafios na triagem e as perspectivas entre identificação dessas variações e a compreensão da influencia nas doenças neuropsiquiátricas. A partir da revisão de literatura deste capítulo, o foco principal do nosso trabalho passou a ser o uso de ferramentas computacionais na busca de novos fatores de risco genético, como será apresentado no capítulo 4.

No capítulo 4 apresentamos novos fatores de risco para a DA, a partir da correlação entre os níveis das proteínas  $A\beta$ -40 e  $A\beta$ -42 no plasma sanguíneo e no fluído cerebrospinal (CSF, do Inglês: *Cerebrospinal fluid*) e 598.821 *SNPs*, provenientes do banco de dados ADNI. Entre os novos fatores de risco temos 5 *SNPs* nas regiões dos genes *TOMM40*, *PAMR1*, *TRIM9* e *CCDC112* e 3 *SNPs* em regiões não codificantes dos cromossomos 2, 14 e X.

Por fim, são apresentadas as conclusões do nosso trabalho.

# Capítulo 1

## 2. Revisão Bibliográfica

## 2.1. Transtornos do Humor

Os Transtornos do Humor (TH) compõem um grupo de doenças heterogêneas caracterizadas por alterações na esfera cognitiva das emoções, sentimentos e motivação; dentre os TH mais prevalentes e mais estudados estão: a Depressão Maior, a Distímia e o Transtorno Bipolar (TB). Os TH são divididos em Transtorno Depressivo Unipolar (TU) e Transtornos Bipolares (TB). A presença de um único episódio maníaco já é suficiente para classificar o TB, diferenciando-o de depressão isoladamente. São diagnosticados a partir de sintomas e sinais característicos, que vão desde uma depressão profunda acompanhada por sintomas psicóticos, passando por estados de depressão moderada e irritabilidade, até estados eufóricos brandos ou de mania, também acompanhado por sintomas psicóticos. Estes transtornos geram elevado impacto na sociedade a nível mundial, afetam a qualidade de vida dos portadores e são responsáveis por elevadas taxas de suicídio (MANUAL DIAGNÓSTICO E ESTATÍSTICO DE TRANSTORNOS MENTAIS, 2002; ORGANIZAÇÃO MUNDIAL DE SAÚDE, 2012).

Embora a maioria dos pacientes com TH receba algum benefício com os tratamentos atualmente disponíveis, um dos principais obstáculos para o desenvolvimento de tratamentos mais eficazes de TH tem sido a compreensão limitada da sua fisiopatologia (RUSH *et al.*, 2006; TRIVED *et al.*, 2006).

A manifestação fenotípica da doença inclui não só mudanças do humor, como também uma gama de distúrbios do sistema cognitivo, motor, endócrino, autônomo e alterações do sono (SANACORA *et al.*, 2008). Estudos demonstram que o déficit de neurotransmissores como a norepinefrina, dopamina, serotonina, ácido gama-aminobutírico (GABA, do Inglês: *Gamma-Aminobutyric Acid*) e glutamato possam contribuir com a fisiopatologia dos TH. Além disso, elevados níveis de glutamato e GABA tem sido observado no córtex cerebral de pacientes com distúrbios do humor, sugerindo que alterações em ambos os sistemas de neurotransmissores podem contribuir com a fisiopatologia dos TH (CHOUDARY *et al.*, 2005; HASLER *et al.*, 2007; LUYKX *et al.*, 2012).

Em relação à sua etiologia, os TH surgem a partir da interação complexa de múltiplos genes e fatores estressores ambientais, aos quais estamos expostos diariamente e ao longo da nossa vida, e dificilmente são explicados por um único componente genético ou ambiental (KAPLAN e SADOCK, 2007).

Alteração nos sistemas de neurotransmissores é uma das hipóteses mais aceitas como alterações fisiológicas dos TH. O sistema monoaminérgico (serotonérgico, noradrenérgico e dopaminérgico) tem recebido maior atenção nos estudos neurobiológicos dos transtornos de humor. Com base na hipótese de disfunção dos sistemas monoaminérgicos cerebrais na doença bipolar, um

dos primeiros genes candidato investigado foi o gene que expressa a proteína tirosina hidroxilase (TYH, do Inglês: *Tyrosine Hydroxylase*) uma enzima que limita a velocidade de síntese das monoaminas. Embora a maioria dos estudos concentre-se no sistema serotoninérgico, mostrando possível desregulação neste sistema, e em menor medida no sistema noradrenérgico, também há evidências de alterações de outros sistemas de neurotransmissores, como dopaminérgico, glutamatérgico e gabaérgico, no envolvimento da fisiopatologia dos TH (BOURNE *et al.*, 1968; PARE *et al.*, 1969; ROY *et al.*, 1989; MEANA *et al.*, 1992; HOLEMANS *et al.*, 1993; BOWDEN *et al.*, 1997; KLIMEK *et al.*, 1997; ORDWAY, 1997; ZHU *et al.*, 1999; MANN *et al.*, 2001; MANJI *et al.*, 2001; BRAMBILLA *et al.*, 2003; ZALSMAN *et al.*, 2004; SANACORA *et al.*, 2008; SEQUEIRA *et al.*, 2009). A Figura 1 mostra os neurotransmissores noradrenalina, serotonina e dopamina envolvidos em algumas funções cognitivas.

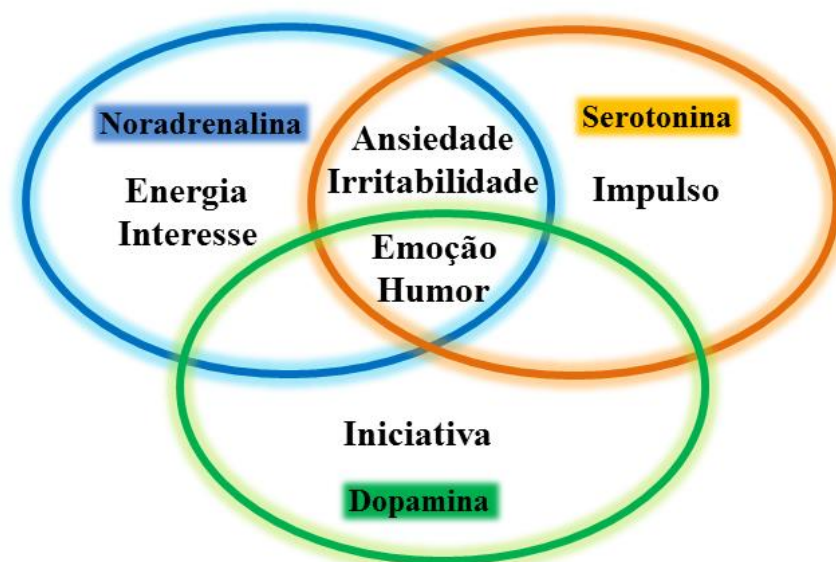


Figura 1. Envolvimento individual e em conjunto dos neurotransmissores noradrenalina, serotonina e dopamina em algumas funções cognitivas.

## 2.2. Fatores de risco genético associado aos TH

Em comparação a outras doenças complexas, as doenças psiquiátricas têm uma herdabilidade relativamente alta (CHAKRAVARTI e LITTLE, 2003). O envolvimento de componentes genéticos em TH é fortemente sugerido, principalmente por estudos envolvendo gêmeos, estudos com famílias e com adotados, além dos estudos de genética molecular, como estudos de ligação e de associação que também apontam para os TH como uma desordem de bases genéticas, envolvendo complexos mecanismos de herança (LIMA *et al.*, 2004; MIDDELDORP *et al.*, 2008).

Ogilvie *et al.* (1996) foram os primeiros a relatar uma associação entre uma repetição em tandem em número variável (VNRT, do Inglês: *Variable Number of Tandem Repeats*) no segundo intron do gene que codifica o transportador da serotonina (5-HTT) e TU. No mesmo ano,

independentemente, Collier *et al.* (1996) não encontraram associação com TU, mas verificaram uma frequência maior de uma variante deste polimorfismo num grupo de pacientes com TB.

Estudos de ligação realizados em pacientes com TB sugeriram o envolvimento de regiões nos cromossomas X, 18, 5, 11, 4, 21 e 12 (por ordem de quantidade de estudos de ligação positivos), que, em geral, contêm genes codificadores de proteínas, ou subunidades de proteínas, relacionadas com neurotransmissores e seu funcionamento (MERIKANGAS *et al.*, 2002).

O Polimorfismo no promotor da proteína transportadora da serotonina (5-HTT) tem sido um dos principais fatores de risco associados aos TH, principalmente em estudos multicêntricos e quando associados a eventos estressores, no entanto, Teste de Desequilíbrio de Ligação (*TDT*, do Inglês: *Transmission Disequilibrium Test*) não têm sido concordantes acerca do envolvimento deste polimorfismo na origem dos TH (OLIVEIRA *et al.*, 2000). Posteriormente, dois polimorfismos nos genes *GRIN1* e *GRIN2B*, que codificam a subunidade NR1 e NR2B respectivamente (subunidades de receptores ionotrópicos ativadas pelo glutamato), foram associados com TB (MARTUCCI *et al.*, 2006).

### **2.3. Genes dos sistemas glutamatérgico e gabaérgico inseridos no estudo relacionado aos TH**

Glutamato e GABA são os principais neurotransmissores excitatório e inibitório no sistema nervoso central, respectivamente (DANBOLT, 2001; SCHOUSBOE e WAAGEPETERSEN, 2007). E o papel dos sistemas glutamatérgico e gabaérgico vem sendo investigado no envolvimento fisiopatológico e no tratamento dos distúrbios de humor, estudos sugerem que alterações em ambos os sistemas de neurotransmissores podem contribuir com tanto com a fisiopatologia dos TH quanto com a resposta ao tratamento (MERALI *et al.*, 2004; CHOUDARY *et al.*, 2005; SANACORA *et al.*, 2008). Pois, alterações nos níveis de glutamato e GABA tem sido observado no córtex cerebral de pacientes com distúrbios do humor (SANACORA *et al.*, 2008; SEQUEIRA, 2009).

Embora estudos levantem hipóteses do envolvimento desses neurotransmissores nos TH, os mecanismos moleculares que podem causar ou contribuir com a doença permanecem desconhecidos. Os locos cromossômicos dos genes inseridos neste estudo estão representados na Figura 2.

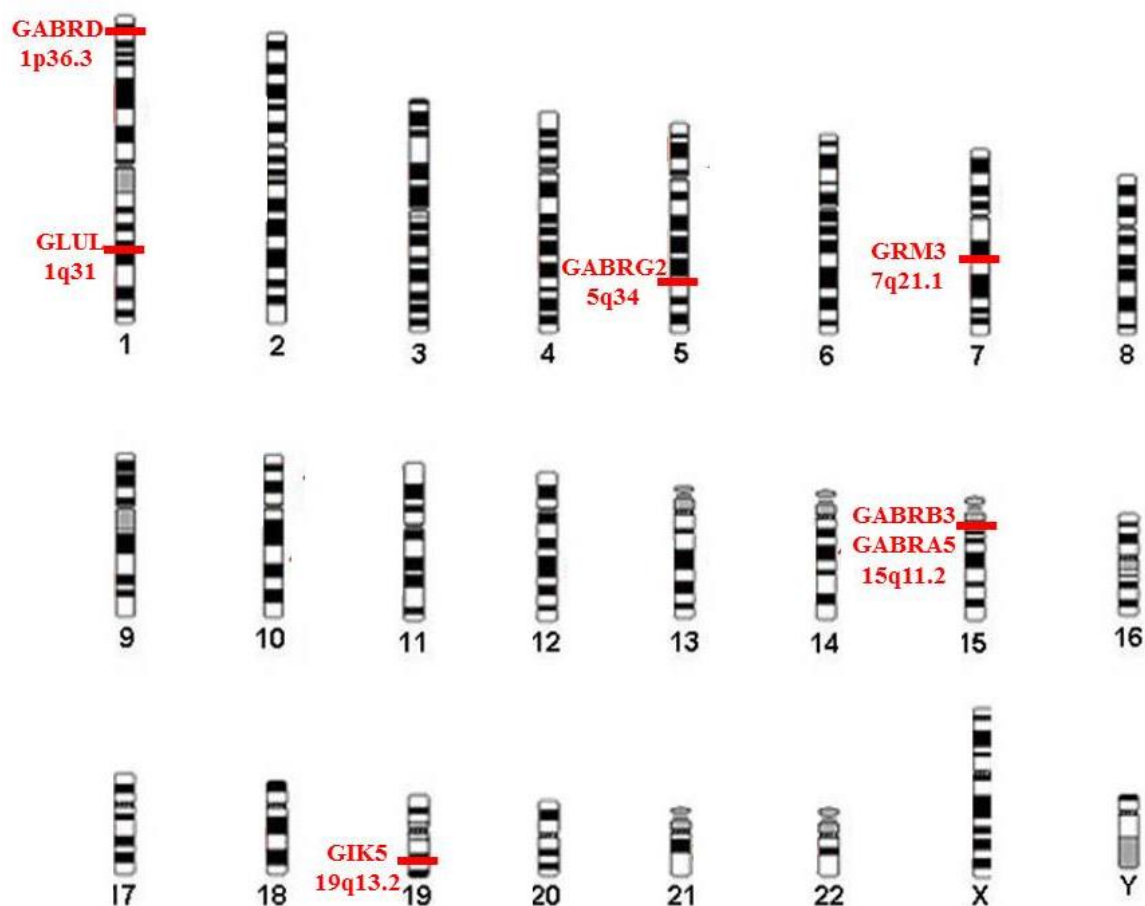


Figura 2. Representação geral dos locos cromossômicos dos genes envolvidos nos sistemas de neurotransmissores glutamatérgico e gabaérgico, que apresentaram níveis de expressão alterada, de acordo com o estudo de *Microarray* de expressão, realizado por Choudary *et.al.*, 2005. Esses genes foram inseridos no nosso estudo de busca por novos potenciais polimorfismos genéticos de risco em TH.

### 2.3.1. Gene *GLUL*

A proteína codificada pelo gene *GLUL* pertence à família glutamina sintetase. O ciclo glutamato-glutamina entre neurônios e astrócitos tem sido proposta como principal via de reciclagem do glutamato, que ao ser recaptado é convertido à glutamina pela ação da glutamina sintetase, que posteriormente é transportada aos neurônios e revertida a glutamato pela ação da glutaminase (DANBOLT, 2001). Localizado na região 1q31, este gene é expresso durante o início da fase fetal, desempenhando um importante papel no controle do pH do corpo através da remoção da amônia em circulação (CLANCY *et al.*, 1996). Em 2009, Sequeira *et al.*, 2005, encontraram nível alterado na expressão de *GLUL* no córtex pré-frontal de suicidas portadores de Depressão Maior.

### 2.3.2. Genes *GRIK5* e *GRM3*

Os receptores de glutamato medeiam a maior parte da neurotransmissão excitatória no cérebro de mamíferos e são ativados em uma variedade de processos neurofisiológicos normais



(SAGER *et al.*, 2009). Esses receptores são classificados em dois grupos ionotrópicos e metabotrópicos, de acordo com suas características moleculares (OZAWA *et al.*, 1998).

Há três classes de receptores ionotrópico de glutamato: NMDA (N-metil-D-aspartato), AMPA (alfa-amino-3-hidroxi-5-metil-4-isoxazol propiônico) e receptores Kainato (YAN e TOMITA, 2012). A Figura 3 ilustra a distribuição desses receptores na membrana neuronal.

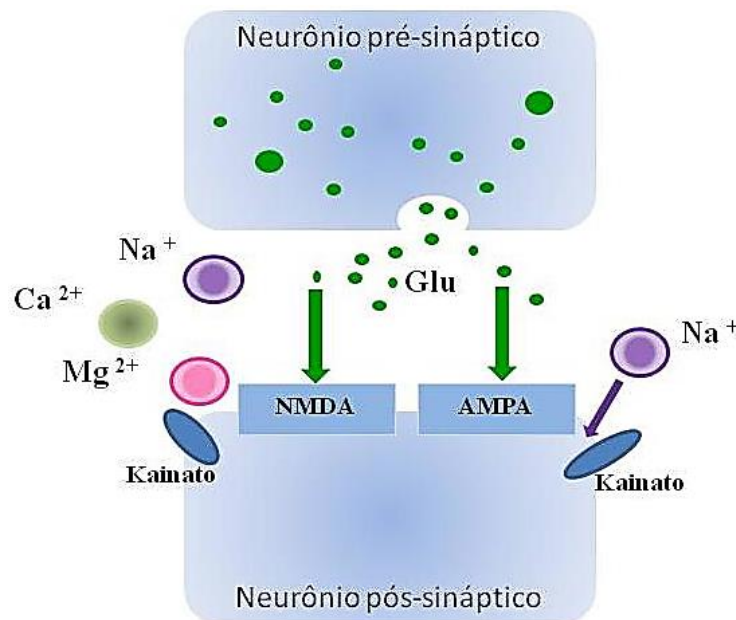


Figura 3. Receptores ionotrópico de glutamato: NMDA, AMPA e Kainato. Distribuídos na membrana neuronal.

O gene *GRIK5*, em humanos localizado na região 19q13.2, codifica o receptor ionotrópico de glutamato, Kainate 5 (SZPIRER *et al.*, 1994). Um recente estudo mostrou possível associação do gene *GRIK5* a diversas desordens neuropsiquiátricas, incluindo os TH (GRATACÒS *et al.*, 2009).

Mapeado no cromossomo 7q21.1, o gene *GRM3* codifica receptores de glutamato metabotrópico 3 (SCHERER *et al.*, 1996). Em estudo de associação, Tsunoka *et al.*, 2009, encontraram uma associação entre um SNP (rs6465084) no gene *GRM3* e Depressão Maior, no entanto, nenhuma associação entre *GRM3* e TB.

### 2.3.3. Genes *GABRB3*, *GABRD*, *GABRA5* e *GABRG2*

Os receptores gabaérgicos são classificados em ionotrópicos (GABA A) e metabotrópicos (GABA B). Consistem em receptores transmembrana, compostos por cinco subunidades organizadas em torno de um canal iônico ( $\text{Cl}^-$ ) central, sendo esse aberto a partir de diferentes ligantes, como o neurotransmissor GABA, benzodiazepínicos, barbitúricos e esteróides anestésicos, entre outros (WEI *et al.*, 2003).

Em humanos, as subunidades dos receptores são divididas nos seguintes tipos: seis tipos de subunidades  $\alpha$  (*GABRA1*, *GABRA2*, *GABRA3*, *GABRA4*, *GABRA5* e *GABRA6*); três tipos de

subunidades  $\beta$  (GABRB1, GABRB2 e GABRB3); três tipos de subunidades  $\gamma$  (GABRG1, GABRG2 e GABRG3); uma subunidade  $\delta$  (GABRD); uma subunidade  $\epsilon$  (GABRE); uma subunidade  $\pi$  (GABRP); e uma subunidade  $\theta$  (GABRQ) (MOKRAB *et al.*, 2007).

Essas subunidades se combinam de diferentes maneiras formando os canais GABA A. Existe um requisito mínimo de duas subunidades  $\alpha$  e  $\beta$  para que seja formado um canal iônico, mas o mais comum é a união de duas subunidades  $\alpha$ , duas  $\beta$  e uma  $\gamma$ , formando o pentâmero  $\alpha_2 \beta_2 \gamma$  (CONNOLLY *et al.*, 1996), representado na Figura 4.

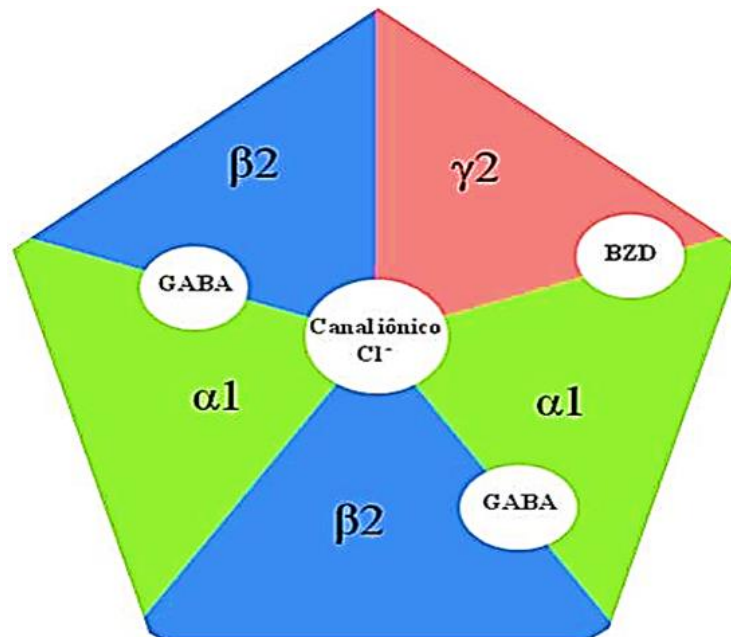


Figura 4. Esquema de um receptor de GABA A formando o pentâmero  $\alpha_2 \beta_2 \gamma$ .

O receptor do ácido gama-aminobutírico subunidade  $\beta$ -3 é uma proteína codificada pelo gene *GABRB3*, que está localizado no cromossomo 15q11.2 (PAPADIMITRIOU *et al.*, 2001). Mutações nesse gene têm sido associadas à patogênese de várias doenças neuropsiquiátricas como autismo, Depressão Maior, esquizofrenia e TB (CRADDOCK *et al.*, 2010; DELAHANTY *et al.*, 2011).

O gene *GABRD* codifica o receptor GABA A subunidade  $\delta$  e encontra-se localizado na região 1p36.3 (EMBERGER *et al.*, 2002) em conjunto com o gene *GABRA5*, da região 15q11.2, que codifica o receptor GABA A  $\alpha$  5, tem sido associado a diversas doenças psiquiátricas. Segundo CRADDOCK *et al.* (2010), variações em ambos os genes apresentam fortes evidências de risco para o TB.

O gene *GABRG2*, localizado na região 5q34, codifica o receptor GABA A subunidades  $\gamma$  2. A região cromossômica 5q34, região onde o receptor GABA A subunidades  $\gamma$  2 é mapeado, foi identificada como uma região de suscetibilidade para os TH (YAMADA *et al.*, 2003).

## 2.4. Demências e Doença de Alzheimer

Demência é um termo genérico, utilizado para definir uma síndrome clínica caracterizada pela deterioração progressiva de múltiplos domínios da cognição, com intensidade suficiente para comprometer a autonomia do indivíduo. Representa umas das principais causas de incapacidade, institucionalização e redução de sobrevida na população, em especial a população idosa (MANUAL DIAGNÓSTICO E ESTATÍSTICO DE TRANSTORNOS MENTAIS, 2002). A demência tem um impacto importante na vida social e na dimensão emocional das famílias e dos seus cuidadores em toda parte no mundo (RADEMAKERSR e ROVELET-LECRUX, 2009).

As síndromes demenciais são classificadas em demências degenerativas, demências não degenerativas e demência mista, além de serem subdivididas em irreversíveis e potencialmente reversíveis. As demências degenerativas englobam: Doença de Alzheimer (DA), Doença de Parkinson, Demência por Corpos de Lewy (DCL) e Demência Fronto-temporal (DFT). Já entre as demências não degenerativas, destacam-se a Demência Vascular (DV), as Demências Hidrocefálicas (DH), as demências por lesões expansivas intracranianas e as Demências Toxicometabólicas (DT). Entre as demências irreversíveis encontram-se todas as demências degenerativas, além da DV. Por outro lado, entre as demências com potencial reversibilidade destacam-se as DH, as demências por lesões expansivas intracranianas e as DT (FORNARI *et al.*, 2010).

As funções cognitivas comumente acometidas nos casos de demência são apresentadas por: distúrbios da fala (afasia), incapacidade de reconhecer ou identificar objetos (agnosia) e perda da habilidade de executar movimentos (apraxia) (BROOKS e BASTOULY, 2004). Os déficits cognitivos geralmente apresentam um declínio em relação a um nível anterior de funcionamento, capaz de comprometer o funcionamento ocupacional ou social do indivíduo, geralmente somados a sintomas neuropsiquiátricos, tais como depressão, ansiedade, alterações do sono e agressividade (ALZHEIMER'S ASSOCIATION, 2013).

A síndrome demencial pode afetar cada indivíduo com severidade diferente, embora normalmente não faça parte do envelhecimento, a demência afeta principalmente as pessoas mais idosas, onde a prevalência da demência aumenta de forma exponencial com a idade, e dobra a cada cinco anos a partir dos 65 anos de idade, chegando à prevalência de 30% aos 85 anos de idade (APRAHAMIAN *et al.*, 2009). Estima-se que cerca de 35,6 milhões de pessoas têm demência, e a cada ano 7,7 milhões de novos casos são registrados devido ao envelhecimento da população, e há uma estimativa de que o número de pessoas com demência dobre a cada 20 anos. De acordo com a organização mundial de saúde, a DA é a causa mais comum de demência, ocorrendo em 40% dos indivíduos com mais de 85 anos de idade e contribuindo com 60-70% dos casos de demência em

todo o mundo, afetando cerca de 26 milhões de pessoas, com estimativa de que até o ano de 2050 cerca de 100 milhões de pessoas desenvolvam a doença (ORGANIZAÇÃO MUNDIAL DE SAÚDE, 2012). Clinicamente, é caracterizada por perda progressiva da memória e da capacidade de raciocínio, mudanças de humor, mudanças de personalidade e perda de independência. Essas doenças resultam dos efeitos combinados de fatores ambientais e de variações genéticas de um número elevado de genes, que podem levar tanto a disfunções sinápticas quanto a degeneração neuronal (DN) (RASCHETTI *et al.*, 2007; BIHAGI *et al.*, 2012).

Os sinais e sintomas clássicos associados a DA incluem alterações do humor (depressão), do comportamento (desinibição e agressividade), das funções cognitivas (insônia, hipersonia e alterações do apetite), da sensopercepção (alucinações e delírios) e sensitivomotoras (mioclonia e sinais extrapiramidais), esses sintomas se iniciam de forma branda e tornam-se mais severos à medida que a doença evolui (SPERLING *et al.*, 2011).

De acordo com a progressão dos sintomas, a DA pode ser dividida em três fases, uma fase pré-sintomática inicial, um estágio prodrômico conhecido como comprometimento cognitivo leve (MCI, do Inglês: *Mild Cognitive Impairment*) e uma terceira fase, quando os pacientes apresentam demência com deficiências em vários domínios e perda de função nas atividades diárias (TROJANOWSKI *et al.*, 2010).

Na fase pré-sintomática da doença, os sintomas apresentados pelo indivíduo, como por exemplo, lapsos de memória, muitas vezes são confundidos com o processo natural do envelhecimento e não há procura de ajuda médica para investigar as origens dos esquecimentos, conseqüentemente, o diagnóstico precoce da doença torna-se difícil (ALZHEIMER'S ASSOCIATION, 2013).

MCI caracteriza uma condição clínica na qual o paciente possui comprometimento cognitivo, mas não apresenta demência, pois desempenha normalmente suas atividades cotidianas (HANSSON *et al.*, 2006). Os pacientes com MCI constituem um grupo heterogêneo, pois alguns pacientes têm apenas a memória afetada, enquanto outros apresentam comprometimento de outras funções como a linguagem, por exemplo. Além de heterogêneo, o grupo também é instável. Há evidências de que indivíduos com MCI, em curto prazo, podem desenvolver quadro demencial da DA, pois, cerca de 80% dos pacientes desenvolvem demência em algum momento das suas vidas, enquanto outros nunca desenvolvem demência propriamente dita (ROSENBERG e LYKETSOS, 2008; SPERLING *et al.*, 2011). MCI está associado ao envelhecimento, havendo prevalência em indivíduos acima de 60 anos de 5% e eleva-se para 15% na faixa etária dos 75 anos (ROSENBERG e LYKETSOS, 2008). Existem diversos marcadores clínicos, tais como sintomas depressivos e déficits cognitivos, e marcadores biológicos como, por exemplo, a redução do nível de  $\beta A$  no CSF (SPERLING *et al.*, 2011; WEINER, *et al.*, 2012).

Há sugestões que as modificações no estilo de vida, incluindo exercícios físicos, atividades de lazer e atividades de estimulação cognitiva podem ser eficazes na prevenção da progressão do MCI (ROSENBERG e LYKETSOS, 2008). Assim, é considerado que MCI pode ser utilizada para elucidar as alterações fisiopatológicas que ocorrem desde a fase pré-sintomática até a fase demencial (SPERLING *et al.*, 2011; WEINER, *et al.*, 2012).

Na fase final, há declínio cognitivo muito grave de todas as funções cognitivas, característico do quadro demencial da DA, o cérebro encolhe drasticamente devido à morte celular generalizada e nessa fase, o paciente requer cuidados em tempo integral, pois perde a capacidade de cuidar de si, de se comunicar e de reconhecer a família e entes queridos (WEINER, *et al.*, 2012).

A DA tem como características neuropatológicas a deposição do peptídeo  $\beta$ -amiloide ( $\beta$ A) no cérebro, formando as placas amiloides ou placas senis, e nos vasos sanguíneos, originando a angiopatia amiloide; e hiperfosforilação da proteína TAU, formando os emaranhados neurofibrilares (MUDHER e LOVESTONE, 2002; TROJANOWSKI *et al.*, 2010). A formação dos agregados do peptídeo  $\beta$ A resultam do metabolismo anormal da proteína percussora amiloide (APP), sendo considerado biomarcador clínico potencial (TARAWNEH e HOLTZMAN, 2010). Por ser detectável no CSF e no plasma sanguíneo, este biomarcador tem sido utilizado para o diagnóstico clínico (KOYAMA *et al.*, 2012). Os emaranhados neurofibrilares intracelulares são formados por desestruturação do citoesqueleto dos microtúbulos, devido à hiperfosforilação da proteína TAU, acompanhada por perda massiva de neurônios. A proteína TAU, pode ser fosforilada em vários sítios, sendo P-tau 181, 199, 231, 235, 396, e 404, os sítios específicos que têm sido estudados e associados a DA (THAL *et al.*, 2006). Os agregados do peptídeo  $\beta$ A e os emaranhados neurofibrilares tendem a se espalhar através do córtex em um padrão previsível, à medida que a doença progride (DICKERSON *et al.*, 2011). A Figura 5 ilustra a progressão da DA no cérebro.

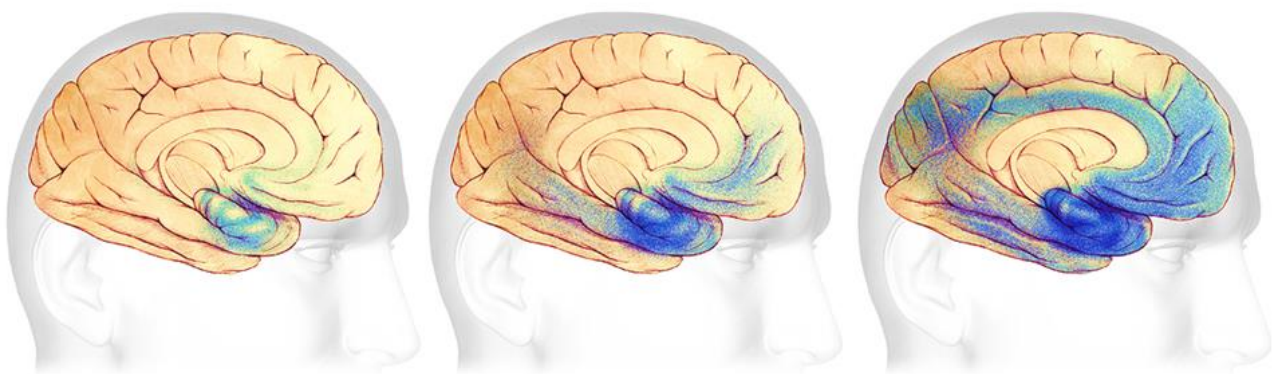


Figura 5. Progressão da DA no cérebro. Placas e emaranhados, representados nas áreas azuis, tendem a se espalhar através do córtex em um padrão previsível, à medida que a DA progride, em uma taxa de progressão é variável entre os pacientes. Adaptado de: <http://www.alz.org/braintour/progression.asp>.

## 2.5. Genética na Doença de Alzheimer

O envolvimento de componentes genéticos na DA é fortemente sugerido. Pesquisas realizadas com gêmeos sugerem a presença de fatores genéticos como responsáveis pela DA, este aspecto também é reforçado devido a gêmeos monozigóticos apresentarem risco cinco vezes maior de desenvolverem a DA quando comparados aos gêmeos dizigóticos (GATZ *et al.*, 2006; GATZ e PEDERSEN, 2013). Adicionalmente, existe uma forte correlação entre mutações autossômicas dominantes nos genes *APP*, *PSEN1* e *PSEN2* atuando de forma mendeliana em casos de DA familiar de início precoce, considerada uma forma rara da doença, por representar apenas 5% dos casos, acometendo indivíduos na faixa etária de 30 a 65 anos de idade (LAMBERT *et al.*, 2009; CRUTS *et al.*, 2012). Já para os casos de Alzheimer onde indivíduos são afetados após os 65 anos de idade e corresponde a maioria dos casos da doença, existe a associação do gene *APOE*  $\epsilon$ 4 com o início tardio da DA.

*APOE*  $\epsilon$ 4 é um dos três alelos ( $\epsilon$ 2,  $\epsilon$ 3 e  $\epsilon$ 4) do gene *APOE*, responsável por produzir uma proteína que transporta o colesterol no sangue. Indivíduos que herdam pelo menos um alelo *APOE*  $\epsilon$ 4 têm um risco aumentado de desenvolver a doença de Alzheimer e de desenvolvê-lo mais cedo do que aqueles que herdam as isoformas  $\epsilon$ 2 ou  $\epsilon$ 3 do gene *APOE*, enquanto aqueles que herdam dois genes *APOE*-  $\epsilon$ 4 têm um risco ainda maior. No entanto, a herança de uma ou duas cópias do gene não é determinate para que o indivíduo desenvolva a DA, pois cerca de 50% dos idosos portadores do alelo  $\epsilon$ 4 do gene *APOE* não chegam a desenvolver demência, fato que corrobora com a busca de novos fatores genéticos de risco para a DA (ALZHEIMER'S ASSOCIATION, 2012).

A busca por fatores genéticos que influenciam na fisiopatologia da DA levou a identificação de uma série de polimorfismos genéticos comumente presentes na população em geral, mas que podem atuar como modificadores da suscetibilidade à doença. Recentemente, através de análises de GWAS foram encontrados locis de suscetibilidade à DA que são comuns na população em geral e que exercem efeitos de risco ao desenvolvimento da doença, mas não o suficiente para causar a doença (BERTRAM *et al.*, 2010; SPERLING *et al.*, 2011). Portanto, a compreensão das bases genéticas envolvidas no mecanismo das doenças complexas, como no caso da DA, tal compreensão fornecerá benefícios tanto no diagnóstico, como no tratamento dessas doenças.

## 2.6. Polimorfismos Genéticos

Polimorfismos são variações na sequência da molécula de DNA, podem ser trocas, repetições, inserções ou deleções de uma ou mais bases nitrogenadas. *SNPs* são variações na sequência de DNA que ocorrem quando um único nucleotídeo da sequência é alterado, levando a uma alternativa usual de dois nucleótidos possíveis numa dada posição. Esse tipo de variação é o mais frequente no genoma humano, estão distribuídos ao longo de todo genoma, ocorrendo em média a cada 5.000 pares de base, sendo as mais comuns variações genéticas, correspondendo a aproximadamente 90% das variações genéticas em humanos (PUSHKAREV *et al.*, 2009; BLENDY, 2011). Como consequência da troca de uma base, podem ocorrer dois tipos de mutação: transição e transversão. Mecanismo de mutação que resulta em transições ocorre troca entre: purina-purina ou pirimidina-pirimidina, enquanto que nas transversões ocorre trocas entre: purina-pirimidina ou pirimidina-purina, não incluindo polimorfismos de inserção ou deleção. Os *SNPs* podem estar presentes em qualquer região do genoma, tais como introns, exons, regiões promotoras e intergênicas, e sua presença pode levar ou não a alteração dos aminoácidos de uma proteína (KIM e MISRA, 2007).

Com a evolução dos métodos de sequenciamento de nova geração, as inserções e deleções, conhecidas como *INDELs*, passaram a corresponder à segunda classe de variação genética mais comum em humanos (KU *et al.*, 2010). Há milhões de *INDELs* descritos, sendo os *INDELs* dialélicos os mais abundantes, representando 8% de todas as variações genéticas encontradas no DNA humano (WEBER *et al.*, 2002; LEVY *et al.*, 2007; MULLANEY *et al.*, 2010). A identificação e compreensão do impacto de *SNPs* e *INDELs* no genoma podem ter um impacto importante de como os seres humanos respondem à doença, ao meio ambiente e a medicamentos, podendo ser utilizados como marcadores genéticos em busca de fatores que influenciam as doenças neuropsiquiátricas (BLENDY, 2011; DEPRISTO *et al.*, 2011).

A triagem desses polimorfismos pode ser realizada por métodos experimentais, tais como: reação em cadeia da polimerase (*PCR*, do Inglês: *Polymerase Chain Reaction*), sequenciamento automático, *PCR* em tempo real, pirosequenciamento, biosensores para DNA e *GWAS*. Outra forma de realizar a triagem de alterações genéticas é por meio do uso de ferramentas computacionais, envolvendo uma grande quantidade de dados que podem indicar potenciais novos polimorfismos, que serão alvos de pesquisas de genotipagem (NAVRATIL *et al.*, 2008; MARIAN, 2012; WAGNER, 2013).

Acredita-se que os *SNPs* e os *INDELs* sejam as mais frequentes classes de mutações que predisõem as doenças neuropsiquiátricas, incluindo TH e DA. Embora nem todos os polimorfismos estejam associados à doença. O grau com que essas variações afetam o fenótipo é o

objetivo maior dos estudos de associação de *SNPs* e *INDELs*. No entanto, a detecção de *SNPs* e de *INDELs* com precisão continua a ser um desafio significativo (GAO *et al.*, 2008; PELLEYMOUNTER, *et al.*, 2011; HAN *et al.*, 2012).

## **2.7. Estudos Genéticos de Associação**

Existem diferentes abordagens para se identificar polimorfismos genéticos associados a determinado fenótipo, classicamente é utilizado estudos de associação genética. Em geral, estudos de associação genética testam dois grupos fundamentalmente distintos para examinar genes candidatos de interesse, escolhidos com base na hipótese de relevância biológica para a doença em estudo. Sendo os estudos caso controle o tipo mais amplo de estudo de associação, neles buscam-se contribuições genéticas que possam auxiliar na caracterização de uma determinada doença, partindo do princípio que alelos que predispõe à doença deverão ser mais frequentes em pacientes comparados com o controle, da mesma forma que os alelos que possam oferecer proteção contra o desenvolvimento de uma condição patológica devem aparecer em maior número no grupo controle (NSENGIMANA e BISHOP, 2012).

Com o avanço tecnológico que possibilitou diminuir o custo do sequenciamento genético, os estudos que comparam as frequências alélicas de todos os marcadores polimórficos disponíveis em indivíduos não relacionados que possuam um determinado sintoma ou condição patológica e em controles saudáveis se disseminaram. Até o final do primeiro semestre de 2012 já haviam sido publicados 1350 GWAS englobando 18 categorias que abrangem diferentes doenças e características humanas, como mostra a Figura 6, disponível em: <http://www.genome.gov/GWAStudies> (HINDORFF *et al.*, 2012).





Figura 6. GWAS publicados a partir de 2005 até junho de 2012, posicionados de acordo com a localização do lócus estudado nos cromossomos humanos; os círculos coloridos representam as categorias das doenças estudadas, os círculos amarelo-claros englobam as doenças neuropsiquiátricas, incluindo a DA. Disponível em: <http://www.genome.gov/GWASStudies> (HINDORFF *et al.*, 2012).

Atualmente, o GWAS é um dos métodos experimentais mais utilizados para estudos de associação onde se busca investigar variações genéticas associadas a doenças complexas. Em um GWAS a população amostral é dividida em casos e controles, sendo o grupo dos casos formado por indivíduos afetados pela doença que se quer investigar e o grupo controle representados por indivíduos saudáveis. Nesses grupos são caracterizadas centenas de milhares de *SNPs*, em geral de 300.000 a 500.000 *SNPs* representativos da população em estudo, possibilitando uma investigação ampla do genoma humano (HINDORFF *et al.*, 2009).

A principal vantagem nesse tipo de estudo é a obtenção de uma fonte muito rica de informações sobre a variabilidade genética humana, permitindo assim o estudo de doenças complexas, como no caso da DA, que se destaca entre uma das mais investigadas através de estudos de GWAS (REIMAN *et al.*, 2007; HAROLD *et al.*, 2009; HINDORFF *et al.*, 2009). No entanto, a análise de dados de GWAS é um problema computacional desafiador, devido à dimensionalidade e complexidade dos dados. Assim, paralelamente ao crescimento dos estudos de investigação do genoma humano, abordagens computacionais que viabilizem a análise dos dados gerados por essas investigações têm sido desenvolvidas (DUAN *et al.*, 2013).

## 2.8. Bioinformática

A bioinformática é uma área interdisciplinar que combina biologia molecular e ciência da computação e emergiu em meados da década de 90 junto com o surgimento dos sequenciadores automáticos de DNA que geraram rapidamente um grande número de sequências gênicas a serem armazenadas e analisadas. Centra-se no desenvolvimento de bancos de dados, algoritmos computacionais de busca, software de previsão e análise de sequências de genes e de proteínas (PIERCE, 2012).

A bioinformática desenvolve e aplica essas ferramentas para minerar os dados e extrair a informação útil a partir de projetos de sequenciamento. Apresenta-se em constante crescimento junto à evolução das técnicas de sequenciamento que englobam todo o genoma e são capazes de gerar informações sobre milhões de pares de bases simultaneamente, que consequentemente trazem novas informações genéticas que requerem o contínuo desenvolvimento de métodos computacionais cada vez mais sofisticados para armazenar, recuperar, comparar e analisar dados genéticos (PIERCE, 2012). Assim, passou a ser considerada por grande número de pesquisadores como ferramenta fundamental para o armazenamento e análise de dados genômicos.

Em análise de dados de *GWAS*, a imensa quantidade de dados se torna um fator crucial na seleção dos métodos mais adequados, já que nem todos os métodos são capazes de lidar com elevado número de dados em um tempo hábil, pois o tempo computacional pode ser significativamente reduzido com o uso de uma técnica apropriada (DUAN *et al.*, 2013).

Os métodos de busca gulosa iniciam pela busca do *SNP* que melhor discrimina casos de controles e seguem selecionando *SNPs* que, ao serem acrescentados ao grupo, reduzem um determinado critério de erro. Os *SNPs* vão sendo selecionados até que se atinja o mínimo global ou determinada condição de parada. O software PLINK destaca-se entre as ferramentas que utiliza esse método de busca (PURCELL *et al.*, 2007).

### 2.8.1. Alinhamento de Sequências

Uma atividade extremamente importante na Bioinformática é o alinhamento ou comparação de sequências biológicas, onde é possível de forma rápida agregar informações sobre uma sequência desconhecida e compará-la com um banco de dados de sequências conhecidas. O alinhamento de sequências pode ser classificado em dois tipos (LESK, 2008):

- Alinhamento global: encontra o melhor alinhamento de uma sequência inteira com outra(s) sequência(s) inteira.

- Alinhamento local: encontra o melhor alinhamento entre algum segmento de uma sequência com algum segmento de outra(s) sequência(s).

O algoritmo de alinhamento global foi aplicado pela primeira vez por S. B. Needleman e C. D. Wunsch, posteriormente T. Smith e M. Waterman modificado para identificar alinhamentos locais (LESK, 2008). Atualmente os algoritmos mais utilizados para pesquisa de similaridade em banco de dados são: Smith-Waterman e o BLAST (BLAST do Inglês: *Basic Local Alignment Search Tool*) (SMITH e WATERMAN, 1981; ALTSCHUL *et al.*, 1990; TATUSOVA e MADDEN, 1999; LESK, 2008).

O algoritmo Smith-Waterman é o método que aperfeiçoa o alinhamento local. Levando em consideração a presença de inserções e/ou deleções, permite o alinhamento tanto de sequências com tamanhos diferentes, quanto de sequências com apenas alguns trechos conservados (SMITH e WATERMAN, 1981).

O algoritmo BLAST utiliza heurística e algoritmos de programação dinâmica. Foi desenvolvido a partir do algoritmo Smith-Waterman, com o objetivo de melhorar o tempo de execução. No entanto, não apresenta a mesma precisão que o Smith-Waterman na análise de sequências pequenas (ALTSCHUL *et al.*, 1990), mas o seu tempo de execução é significativamente menor.

Nos estudos biológicos, é fundamental que os alinhamentos indiquem uma real similaridade entre as sequências analisadas. Na análise dos alinhamentos alguns valores são fornecidos como resultado para cada alinhamento gerado. Neste trabalho, utilizamos os valores da identidade e do *e-value* como filtros de seleção das sequências. A porcentagem de identidade quantifica o grau de similaridade entre as sequências comparadas. Enquanto o *e-value* corresponde à probabilidade de se obter, com outra sequência aleatória de mesmo tamanho e composição de nucleotídeos, outro alinhamento com *score* igual ou superior. Isso ocorre devido ao processo de alinhamento ser de forma aleatória, e quando o algoritmo compara fragmentos com diferentes tamanhos, sequências que não têm relação alguma podem ser alinhadas. Desta forma, o *e-value*, mais confiável tendem a zero (ALTSCHUL, 1990). No presente trabalho foram selecionadas sequências que apresentaram identidade entre 97% e 99% e *e-value*  $> 10^{-20}$ .

### 2.8.2. Redes Neurais Artificiais e Software PMut

As redes neurais artificiais (RNAs) são modelos matemático-computacionais desenvolvidos para solucionar problemas através da simulação do cérebro humano. As redes neurais possuem nós ou unidades de processamento simples que calculam determinadas funções matemáticas. Essas unidades equivalem aos neurônios, que se encontram dispostas em uma ou mais camadas e são interligadas por um grande número de conexões, nas quais recebem e enviam informações. As camadas são classificadas em três grupos: a camada de entrada, em que as unidades recebem a

informação que será processada, a camada(s) intermediária(s), onde é feito processamento das informações, e a camada de saída, que conclui e apresenta o resultado final (BRAGA, *et al.*, 2000). A Figura 7 apresenta uma arquitetura de uma RNA com os três grupos de camadas.

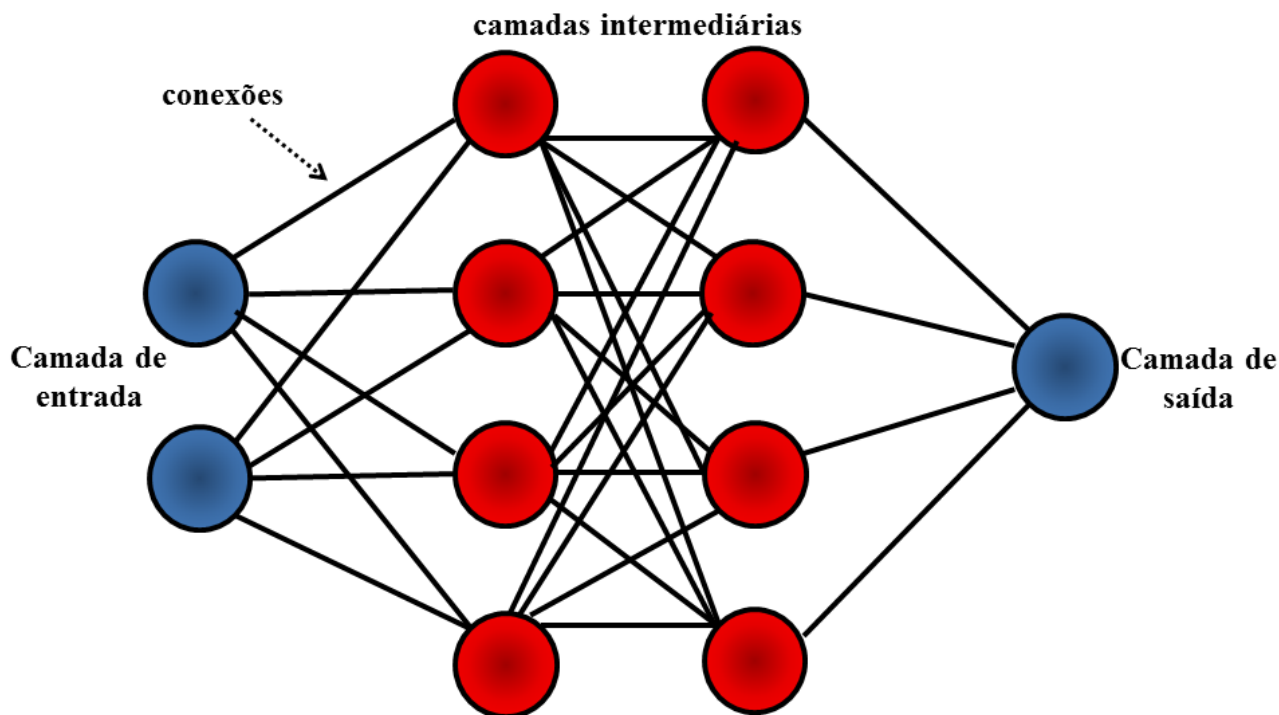


Figura 7. Arquitetura de uma Rede Neural Artificial com os três grupos de camadas.

As RNAs podem ser aplicadas para resolver uma grande quantidade de problemas. Com relação à biologia molecular, um bom exemplo de aplicação são softwares de predição de mutações associadas à doença, como o software PMut (FERRER-COSTA *et al.*, 2005).

PMut é um software disponível na internet (<http://mmb2.pcb.ub.es:8080/PMut/>) que visa à predição rápida e precisa da associação de um polimorfismo de aminoácido único (SAP, do Inglês: *Single Aminoacid Polymorphism*) e um caráter patológico. Em especial para responder à seguinte pergunta: Dada uma mutação em um local específico em uma sequência de aminoácidos, podemos dizer se ela pode ser patológica (isto é, que pode levar a doença para o portador) ou neutra (sem efeito sobre a saúde do portador) (FERRER-COSTA *et al.*, 2005).

Este software é baseado no uso de RNA, treinada com um banco de dados de mutações neutras e mutações patológicas. A taxa de acerto é de 83,5% (FERRER-COSTA *et al.*, 2005), sendo considerado um valor vantajoso quando comparado com os valores de outros programas que realizam a mesma função, como o SIFT (NG e HENIKOF, 2003) e POLYPHEN (SUNYAEV *et al.*, 2001) que exibem uma taxa de acerto de 70,3% e 73,5%, respectivamente. Além disso, o PMut também possui um índice de confiabilidade (variando entre 0 e 9) que pode ser utilizado para

manter apenas as melhores previsões (aqueles com índice de confiabilidade acima de 5) (FERRER-COSTA *et al.*, 2004).

### **2.8.3. Banco de dados Alzheimer's Disease Neuroimaging Initiative – ADNI**

O ADNI foi lançado em 2003, como resultado do consórcio entre instituições acadêmicas, de empresas públicas, de empresas privadas e de instituições sem fins lucrativos, com o objetivo de desenvolvimento de pesquisas clínicas, estudos genéticos, estudos de neuroimagem, estudos bioquímicos e identificação de biomarcadores que auxiliem a detecção precoce e o acompanhamento da DA (WEINER *et al.*, 2012).

Inicialmente, o ADNI, chamado de primeira fase ou ADNI-1, foi projetado para ter duração de cinco anos, durante esse período foram realizadas coletas de informações em série, a cada seis meses, relacionadas ao desempenho cognitivo, mudanças cerebrais estruturais e metabólicas, alterações bioquímicos no sangue, no CSF e na urina do grupo participante do projeto (MUELLER *et al.*, 2005; FRISONI E WEINER, 2010). Fazendo-se cumprir o objetivo primário do ADNI, que foi testar se a ressonância magnética (*MRI*, do Inglês: *Magnetic Resonance Imaging*), tomografia por emissão de pósitrons (*PET*, do Inglês: *Positron Emission Tomography*), biomarcadores, avaliação clínica e neuropsicológica podem ser combinados para medir a progressão do MCI e início da DA; identificando marcadores sensíveis e específicos da progressão AD destinando-se a ajudar pesquisadores e clínicos no desenvolvimento de novos tratamentos e monitorar a sua eficácia, bem como reduzir o tempo e os custos dos testes clínicos (WEINER *et al.*, 2012).

Na fase do ADNI-1, foram acompanhados 800 adultos, com idades entre 55 a 90 anos, pertencentes a mais de cinquenta locais em todo Estados Unidos da América e Canadá. Entre estes 400 indivíduos com MCI (acompanhados por três anos), 200 indivíduos com DA (acompanhados por dois anos), e 200 controles normais (acompanhados por três anos) (WEINER *et al.*, 2012).

Em 2009, o ADNI-1 foi estendido para o ADNI-GO (do Inglês: *Grand Opportunities grant*), onde foi adicionado um novo grupo de 200 pacientes com MCI, para estudo da relação entre biomarcadores em uma fase anterior a progressão da doença. O consórcio foi premiado com um financiamento de 69 milhões de dólares que permitiu, juntamente com o financiamento da indústria farmacêutica, um acordo de cooperação semelhante ao original, ampliando os estudos iniciais. Esta fase é chamada de ADNI-2 e já foram acrescentados 250 pacientes com MCI, 150 novos pacientes com AD e 150 idosos saudáveis (CARRILLO *et al.*, 2009; FRISONI E WEINER, 2010). Prevê-se que o estudo de idosos saudáveis e pacientes com MCI e AD no ADNI-GO e ADNI-2 possam ajudar a identificar indivíduos que são candidatos de risco e a desenvolver terapia preventiva em pacientes quando estão levemente sintomáticos ou assintomáticos (SCHMIDT *et al.*, 2010).

Os dados dos participantes do ADNI encontram-se disponíveis para download no banco de dados no site do ADNI (<http://adni.loni.ucla.edu/>). Nas análises deste trabalho, foram utilizados os participantes da fase ADNI-1. A Figura 8 apresenta o site do banco de dados ADNI.



Figura 8. Página inicial de site do banco de dados ADNI.

#### 2.8.4. PLINK

O PLINK contém um conjunto de ferramentas para análise de varredura genômica e foi projetado para executar diferentes bases de dados simultaneamente. É possível encontrá-lo no link: <http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml>. A Figura 9 apresenta o site do software PLINK.



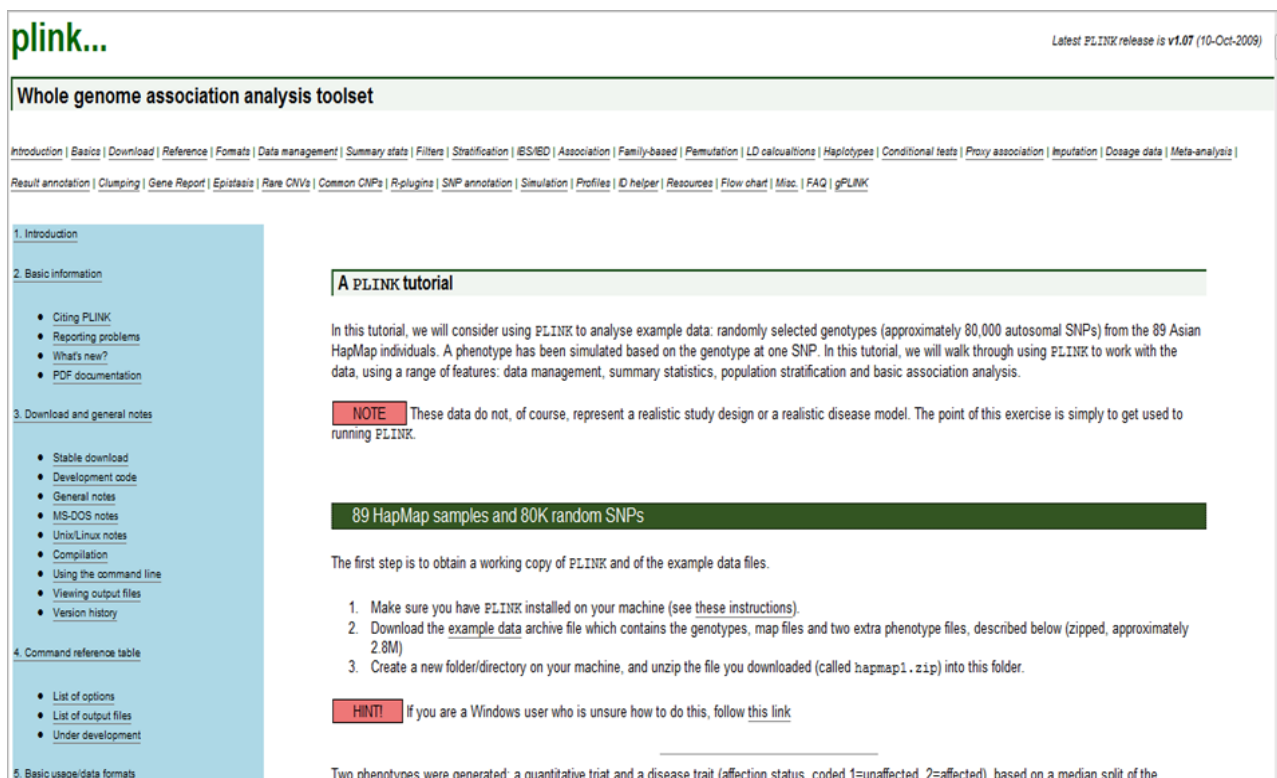


Figura 9. Página inicial do site do software PLINK.

O PLINK interage com um conjunto de ferramentas que realiza análises de associação entre dados de genótipos e fenótipos de maneira computacionalmente eficiente. Devido ao elevado número de dados processados, é sugerido o uso do pacote R para visualizar do arquivo de saída. A linguagem de programação R, altamente difundida na comunidade acadêmica das ciências estatísticas, possui um pacote “*genetics*” de simples uso para análise de dados genéticos (PURCELL *et al.*, 2007).

O PLINK é um programa de linha de comando escrito na linguagem C/C++, tem seu foco puramente na análise de dados de genótipo / fenótipo. Desenvolvido por Shaun Purcell, no Center for Human Genetic Research (CHGR), Massachusetts General Hospital, (MGH), com o apoio da equipe de desenvolvedores do próprio MGH O programa foi desenvolvido sob Licença Pública Geral (GPL), portanto, seu código fonte é de uso livre e encontra-se disponível no seguinte endereço eletrônico: <http://pngu.mgh.harvard.edu/~purcell/plink>. Pode ser executado em diversos sistemas operacionais (PURCELL *et al.*, 2007), neste trabalho foi utilizado o sistema operacional Linux.

Todos os comandos que envolvem PLINK são digitados no *prompt* de comando do sistema operacional, seguido por uma série de opções, de acordo com o teste aplicado (todas as opções iniciando com o caracter -) para especificar os arquivos de dados e/ou métodos a serem utilizados. Todos os resultados são gravados em arquivos com extensões diferentes. Uma lista completa de

todas as opções e tipos de arquivo de saída são dados no endereço eletrônico do próprio programa:  
<http://pngu.mgh.harvard.edu/~purcell/plink>

### 2.8.5. GeneMANIA

GeneMANIA é uma ferramenta de geração hipótese que utiliza grandes conjuntos de dados biológicos disponíveis em banco de dados públicos para encontrar genes correlacionados. Desenvolvido na Universidade de Toronto, no Donnelly Centre for Cellular and Biomolecular Research, nos laboratórios Gary Bader e Quaid Morris. Inicialmente GeneMANIA foi financiado pelo Ontario Genomics Institute, atualmente o financiamento é proveniente do Ontario Ministry of Research and Innovation. Encontra-se disponível gratuitamente através do endereço eletrônico <http://www.genemania.org> e apresentado na Figura 10 (WARDE-FARLEY *et al.*, 2010).

O objetivo do algoritmo do GeneMANIA é possibilitar ao usuário estender uma lista de determinados genes, dados como entrada, para encontrar novos genes que apresentem relação entre si. A partir do conjunto de genes dado pelo usuário é gerada uma rede preditiva de associação que incluem interação proteína-proteína, interação proteína-DNA, vias metabólicas, co-expressão, co-localização, domínio proteico conservado, ortologia e informações sobre doenças.

O GeneMANIA consiste de dois estágios: um algoritmo para cálculo de uma rede de genes com base na função derivado de fontes de dados distintas, e um algoritmo de propagação de rótulos para predição da função gênica dada a rede. A construção da rede baseia-se em um conjunto de dados coletados a partir de 10 bancos de dados públicos e de 6 organismos diferentes. Atualmente a consulta do GeneMANIA abrange 6 organismos (*Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, *Saccharomyces cerevisiae* e *Homo sapiens*) e centenas de conjuntos de dados coletados a partir dos bancos de dados públicos GEO, BioGRID, InterPro, SMART, Pfam, PathwayCommons, BioCyc, Reactome, I2D e Ensembl, que são regularmente atualizados, além de ser indicado um peso que mede cada associação encontrada, dando segurança e precisão nas redes preditas pelo GeneMANIA (MOSTAFAVI *et al.*, 2008; WARDE-FARLEY *et al.*, 2010).



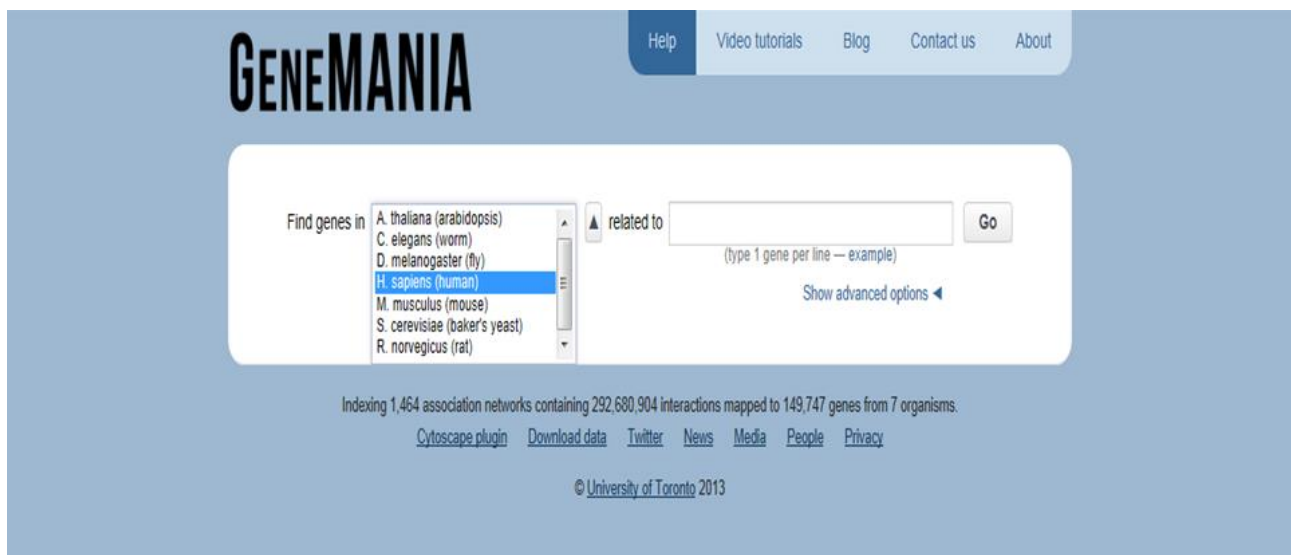


Figura 10. Página inicial do site do software GeneMANIA.

### 3. Referências Bibliográficas

ALTSCHUL, S. F. et al. Basic local alignment tool. *Journal of Molecular Biology*, v. 215, p.403-410, 1990.

ALZHEIMER'S ASSOCIATION. 2013 Alzheimer's disease facts and figures. *Alzheimers Dement.* v. 9, p. 208-245, 2013.

ALZHEIMER'S ASSOCIATION. Alzheimer's disease facts and figures. *Alzheimers Dement.* v. 8, p. 131-168, 2012.

APRAHAMIAN, I.; MARTINELLI, J. E.; YASSUDA, M. S. Doença de Alzheimer: revisão da epidemiologia e diagnóstico. *Rev. Bras. Clin. Med.* v.7, p. 27-35, 2009.

ASSOCIAÇÃO PSIQUIÁTRICA AMERICANA. Manual diagnóstico e estatístico de transtornos mentais. Porto Alegre: Artes Médicas, 2002.

BERTRAM, L.; LILL, C. M.; TANZI, R. E. The genetics of Alzheimer disease: back to the future. *Neuron.* v. 68, p. 270-81, 2010.

BIHAGI, S. W. et al. Do epigenetic pathways initiate late onset Alzheimer disease (LOAD): towards a new paradigm. *Curr Alzheimer Res.* v. 9, p. 574-588, 2012.

BLENDY, J. A. Modeling neuropsychiatric disease-relevant human SNP in mice. *Neuropsychopharmacology.* v. 36, p. 364-365, 2011.

BOURNE, H. R. et al. Noradrenaline, 5-hydroxytryptamine, and 5-hydroxyindoleacetic acid in hindbrains of suicidal patients. *Lancet.* v.2, p. 805–808, 1968.

BOWDEN, C. et al. Dopamine D1 and D2 receptor binding sites in brain samples from depressed suicides and controls. *Brain Res.* v.752, p. 227–233, 1997.

BRAGA, A. P.; CARVALHO A. C. P. L. F.; LUDERMIR T. B. Redes Neurais Artificiais: teoria e aplicações. Livros Técnicos e Científicos (LTC), 2000.

BRAMBILLA, P. et al. GABAergic dysfunction in mood disorders. *Mol Psychiatry*. v. 8, p. 721–737, 2003.

BROOKS, J. B. B.; BASTOULY, V. “Doença de Alzheimer”: uma visão histórica, genética, clínica e terapêutica. *Revista Médica Hospital Ana Costa*, v. 9, p. 52-57, 2004.

CARRILLO, M. C.; SANDERS, C. A.; KATZ, R. G. Maximizing the Alzheimer’s Disease Neuroimaging Initiative II. *Alzheimer’s Dement*. v.5, p. 271–275, 2009.

CHAKRAVARTI, A.; LITTLE, P. Nature, nurture and human disease. *Nature* v. 421, p. 412-414, 2003.

CHOUDARY, P.V. et al. Altered cortical glutamatergic and GABAergic signal transmission with glial involvement in depression. *Proc Natl Acad Sci U S A*. v. 102, p. 15653–15658, 2005.

CLANCY, K. P. et al. Localization of the L-glutamine synthetase gene to chromosome 1q23. *Genomics*. v. 38, p. 418-420, 1996.

COLLIER, D.A. et al. Susceptibility to Bipolar Affective Disorder and Unipolar Depression is Influenced by Allelic Variation of Functional Serotonin Transporter Expression. *Mol. Psychiatry*. v. 1, p. 453-460, 1996.

CONNOLLY, C. N. et al. Assembly and cell surface expression of heteromeric and homomeric gamma-aminobutyric acid type A receptors. *J Biol Chem*. v.271, p.89-96, 1996.

CRADDOCK, N. et al. Strong genetic evidence for a selective influence of GABAA receptors on a component of the bipolar disorder phenotype. *Mol Psychiatry*. v.15, p.146-153, 2010.

CRUTS, M.; THEUNS, J.; VAN BROECKHOVEN, C. Locus-specific mutation databases for neurodegenerative brain diseases. *Hum Mutat*. v. 33, p. 1340-1344, 2012.

DANBOLT, N. C. The glutamate uptake. *Prog. In Neurobiol*. v.65, p. 1-105, 2001.

DELAHANTY, R. J. et al. Maternal transmission of a rare GABRB3 signal peptide variant is associated with autism. *Mol Psychiatry*. v. 16, p. 86-96, 2011.

DEPRISTO, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* v. 43, p. 491–498, 2011.

DICKERSON, B. C. et al. Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology.* v. 76, p.1395-1402, 2011.

DUAN, Q. et al. A comprehensive SNP and indel imputability database. *Bioinformatics.* v. 29, p.:528-531, 2013.

EMBERGER, W. et al. Assignment of the human GABAA receptor delta-subunit gene (GABRD) to chromosome band 1p36.3 distal to marker NIB1364 by radiation hybrid mapping. *Cytogenic and Genome Research.* v. 89, p. 281-282, 2002.

FERRER-COSTA, C.; OROZCO, M.; DE LACRUZ, X. Sequence-based prediction of pathological mutations. *Proteins.* v. 57, p. 811-819, 2004.

FERRER-COSTA, C.; OROZCO, M.; DE LACRUZ, X. Use of bioinformatics tools for the annotation of disease-associated mutations in animal models. *Proteins.* v.61, n.4, p. 878-887, 2005.

FORNARI, L. H. T.; GARCIA, L. P.; HILBIG, A. As diversas faces da síndrome demencial: como diagnosticar clinicamente? *Scientia Medica.* v. 20, p. 185-193, 2010.

FRISONI, G. B.; WEINER, M. W. Alzheimer's disease neuroimaging initiative special issue. *Neurobiol Aging.* v. 31, p.1259–1262, 2010.

GAO, X.; STARMER, J.; MARTIN, E. R. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol.* v. 32, p. 361-369, 2008.

GATZ, M. et al. Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry.* v. 63, p. 168-174, 2006.

GATZ, M.; PEDERSEN, N. L. Study of dementia in Swedish twins. *Twin Res Hum Genet.* v.16, p. 13-16, 2013.

Genetics                      Home                      Reference.                      Disponível                      em:  
<<http://ghr.nlm.nih.gov/handbook/genomicresearch/snp>>. Acesso em: 13/12/2012.

GRATACÒS, M. et al. Identification of new putative susceptibility genes for several psychiatric disorders by association analysis of regulatory and non-synonymous SNPs of 306 genes involved in neurotransmission and neurodevelopment. *Am J Med Genet B Neuropsychiatr Genet.* v.150, p.808-816, 2009.

HAN B. et al. Genetic studies of complex human diseases: characterizing SNP-disease associations using Bayesian networks. *BMC Syst Biol.* v. 6 Suppl 3:S14, 2012.

HANSSON, O. et al. Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurol.* v. 5, p. 228-234, 2006.

HAROLD, D. et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet.* v.41, p. 1088-1093, 2009.

HASLER, G. et al. Reduced prefrontal glutamate/glutamine and gamma-aminobutyric acid levels in major depression determined using proton magnetic resonance spectroscopy. *Arch Gen Psychiatry.* v. 64, p. 193-200, 2007.

HINDORFF, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA.* v.106, p. 9362-9367, 2009.

HINDORFF, L. A., JUNKINS, H. A., MANOLIO, T. A. NHGRI Catalog of published genome-wide association studies. Disponível em: <http://www.genome.gov/GWASStudies>. Acessado em 03/12/2012.

HOLEMANS, S. et al. NMDA glutamatergic receptors, labelled with [3H]MK-801, in brain samples from drug-free depressed suicides. *Brain Res.* v.616, p. 138-143, 1993.

KAPLAN, H. I; SADOCK, B. J. Compêndio de Psiquiatria. 9ª ed. Porto Alegre: Art médica, 2007.

KIM, S.; MISRA, A. SNP Genotyping: Technologies and Biomedical Applications. *Annu. Rev. Biomed. Eng.* v. 9, p. 289-320, 2007.

KLIMEK, V. et al. Reduced levels of norepinephrine transporters in the locus coeruleus in major depression. *J Neurosci.* v.17, p.8451–8458, 1997.

KOYAMA, A. et al. Plasma amyloid- $\beta$  as a predictor of dementia and cognitive decline: a systematic review and meta-analysis. *Arch Neurol.* v. 69, p. 824-831, 2012.

KU, C. S. et al. The discovery of human genetic variations and their use as disease markers: past, present and future. *J Hum Genet.* v. 55, p. 403-415, 2010.

LAMBERT, J. C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nature Genetics.* v. 41, p. 1094-1099, 2009.

LESK, A. M. Introdução à bioinformática. In: Organização e Evolução de Genomas. 2 ed. Porto Alegre: Artmed, 2008. cap. 2, p.87 -132.

LEVY, S. et al. The diploid genome sequence of an individual human. *PLoS Biol.* v. 5(10):e254, 2007.

LIMA, I. V. M.; SOUGEY, E. B.; VALLADA FILHO, H. P. Genética dos transtornos afetivos. *Revista Psiquiatria Clínica.* v.31, p. 34-39, 2004.

LUYKX, J. J. et al. Region and state specific glutamate downregulation in major depressive disorder: a meta-analysis of (1)H-MRS findings. *Neurosci Biobehav Rev.* v. 36, p. 198-205, 2012.

MANJI, H. K.; DREVETS, W. C.; CHARNEY, D. S. The cellular neurobiology of depression. *Nat Med.* v.7, p. 541–547, 2001.

MANN, J. J.; BRENT, D. A.; ARANGO, V. The neurobiology and genetics of suicide and attempted suicide: a focus on the serotonergic system. *Neuropsychopharmacol.* v. 24, p. 467–477, 2001.

MARIAN, A. J. Molecular genetic studies of complex phenotypes. *Transl Res.* v. 159, p. 64-79, 2012.

MARTUCCI, L. et al. n-methyl-D-aspartate receptor NR2B subunit gene GRIN2B in schizophrenia and bipolar disorder: polymorphisms and RNA levels. *Schizophr Res.* v. 84, p. 214-221, 2006.

MEANA J. J.; BARTUREN F.; GARCIA-SEVILLA, J. A. Alpha 2-adrenoceptors in the brain of suicide victims: increased receptor density associated with major depression. *Biol Psychiatry.* v.31, p. 471–490, 1992.

MERALI, Z. et al. Dysregulation in the suicide brain: mRNA expression of corticotropin-releasing hormone receptors and GABA(A) receptor subunits in frontal cortical brain region. *J. Neurosci.* v. 24, p. 1478–1485, 2004.

MERIKANGAS, K. R. et al. Future of genetics of mood disorders research *Biol Psychiatry.* v. 52, p. 457-77, 2002.

MIDDELDORP, C. M. et al. I. Life events, anxious depression and personality: a prospective and genetic study. *Psychological Medicine.* v. 38, p. 1557–1565, 2008.

MOKRAB, Y. et al. Exploring ligand recognition and ion flow in comparative models of the human GABA type A receptor. *J Mol Graph Model.* v. 26, p. 760-774, 2007.

MOSTAFAVI, S. et al. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* v. 9, (Suppl 1):S4, 2008.

MUDHER, A.; LOVESTONE, S. Alzheimer's disease – do tauists and Baptists finally shake hands? *Trend Neurosci.* v. 25, p. 22-25, 2002.

MUELLER, S. G. et al. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am.* v. 15, p. 869–877, 2005.

MULLANEY, J. M. et al. Small insertions and deletions (INDELs). *Hum Mol Genet.* v. 19, p. 131–136, 2010.

NAVRATIL, V. et al. DigiPINS: A database for vertebrate exonic single nucleotide polymorphisms and its application to cancer association studies. *Biochimie*. v. 90, p.563-569, 2008.

NG, P. C.; HENIKOFF, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. v. 31, p.3812-3814, 2003.

NSENGIMANA, J.; BISHOP, D. T. Design considerations for genetic linkage and association studies. *Methods Mol Biol*. v. 850, p. 237-262, 2012.

OGILVIE, A. D. et al. Polymorphism in Serotonin Transporter Gene Associated with Susceptibility to Major Depression. *Lancet*. v. 347, p. 731-733, 1996.

OLIVEIRA, J. R. et al. Analysis of the serotonin transporter polymorphism (5-HTTLPR) in Brazilian patients affected by dysthymia, major depression and bipolar disorder. *Mol Psychiatry*. v. 4, p. 348-349, 2000.

ORDWAY, G. A. Pathophysiology of the locus coeruleus in suicide. *Ann N Y Acad Sci*. v.836, p. 233-252, 1997.

ORGANIZAÇÃO MUNDIAL DE SAÚDE. Disponível em: [http://www.who.int/mental\\_health/management/depression/en](http://www.who.int/mental_health/management/depression/en) Acesso em: 05 dez. 2012.

OZAWA, S.; KAMIYA, H.; TSUZUKI, K. Glutamate receptors in the mammalian central nervous system. *Prog Neurobiol*. v. 54, p. 581-618, 1998.

PAPADIMITRIOU, G. N. et al. GABA-A receptor beta3 and alpha5 subunit gene cluster on chromosome 15q11-q13 and bipolar disorder: a genetic association study. *Am J Med Genet*. v.105, p.317-320, 2001.

PARE, C. M. B. et al. 5-hydroxytryptamine, noradrenaline, and dopamine in brainstem, hypothalamus, and caudate nucleus of controls and patients committing suicide by coal-gas poisoning. *Lancet*. v. 2, p. 133-135, 1969.



PELLEYMOUNTER, L. L. A novel application of pattern recognition for accurate SNP and INDEL discovery from high-throughput data: targeted resequencing of the glucocorticoid receptor co-chaperone FKBP5 in a Caucasian population. *Mol Genet Metab.* v. 104, p. 457-469, 2011.

PIERCE, Benjamin A. Genetics: A Conceptual Approach. 4. ed. New York: W. H. Freeman And Company, 2012. Cap. 20, p. 558-590.

PURCELL, S. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* v. 81, p. 559-575, 2007.

PUSHKAREV, D.; NEFF, N. F.; QUAKE, S. Single-molecule sequencing of an individual human genome. *Nat Biotechnol* v. 27, p. 847–852, 2009.

RADEMAKERS, R.; ROVELET-LECRUX, A. Recent insights into the molecular genetics of dementia. *Trends Neurosci.* v. 32, p. 451-461, 2009.

RASCHETTI, R. et al. Cholinesterase inhibitors in mild cognitive impairment: a systematic review of randomised trials. *PLoS Med.* 4(11):e338, 2007.

REIMAN, E. M. et al. GAB2 alleles modify Alzheimer's risk in APOE carriers. *Neuron.* v. 54, p.713-720, 2007.

ROSENBERG, P. B.; LYKETSOS, C. Mild cognitive impairment: searching for the prodrome of Alzheimer's disease. *World Psychiatry.* v. 7, p. 72-78, 2008.

ROY, A. et al. Suicidal behavior in depression: relationship to noradrenergic function. *Biol Psychiatry.* v.25, p. 341–350, 1989.

RUSH, A. J. et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment stepeds: a STAR\*D report. *Am. J. Psychiatry,* v. 163, p.1905-1971, 2006.

SAGER, C. et al. C-terminal domains of transmembrane alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA) receptor regulatory proteins not only facilitate trafficking but are major modulators of AMPA receptor function. *J Biol Chem.* v.284, p. 32413-32424, 2009.

SANACORA, G. et al. Targeting the glutamatergic system to develop novel, improved therapeutics for mood disorders. *Nat Rev Drug Discov.* v. 7, p. 426-437, 2008.

SCHERER, S.W. et al. Localization of two metabotropic glutamate receptor genes, GRM3 and GRM8, to human chromosome 7q. *Genomics.* v. 31, n. 2, p. 230-233, 1996.

SCHMIDT, M. E. et al. The Alzheimer's Disease Neuroimaging Initiative: perspectives of the Industry Scientific Advisory Board. *Alzheimer's Dement.* v. 6, p. 286–290, 2010.

SCHOUSBOE, A.; WAAGEPETERSEN H.S. GABA: homeostatic and pharmacological aspects. *Prog Brain Res.* v.160, p. 9-19, 2007.

SCHOUSBOE, A.; WAAGEPETERSEN H.S. GABA: homeostatic and pharmacological aspects. *Prog Brain Res.* v.160, p. 9-19, 2007.

SEQUEIRA, A. et al. Global brain gene expression analysis links glutamatergic and gabaergic alterations to suicide and major depression. *PLoS One.* v. 4, p. 65-85, 2009.

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

SPERLING, R. A. et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dementia.* v.7, p.280–292, 2011.

SUNYAEV, S. et al. Prediction of deleterious human alleles. *Hum Mol Genet* v.10, p. 591-597, 2001.

SZPIRER, C. et al. The genes encoding the glutamate receptor subunits KA1 and KA2 (GRIK4 and GRIK5) are located on separate chromosomes in human, mouse, and rat. *Proc Natl Acad Sci U S A.* v.91, p.11849-11853, 1994.

TARAWNEH, R.; HOLTZMAN, D. M. Biomarkers in translational research of Alzheimer's disease. *Neuropharmacology.* v. 59, p. 310-322, 2010.

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.

THAL, L. J. et al. The role of biomarkers in clinical trials for Alzheimer disease. *Alzheimer Dis Assoc Disord*. v. 20, p. 6-15, 2006.

TRIVED, M. H. et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR\*D: implications for clinical practice. *Am. J. Psychiatry*, v.163, p. 28-40, 2006.

TROJANOWSKI, J. Q. et al. Alzheimer's Disease Neuroimaging Initiative. 2010. Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects. *Alzheimer's Dement*. v. 6, p. 230-238, 2010.

TSUNOKA, T. et al. Association analysis of Group II metabotropic glutamate receptor genes (GRM2 and GRM3) with mood disorders and fluvoxamine response in a Japanese population. *Prog Neuropsychopharmacol Biol Psychiatry*. v. 33, p. 875-879, 2009.  
v. 71, p. 854–862, 2002.

WAGNER, M. J. Rare-variant genome-wide association studies: a new frontier in genetic analysis of complex traits. *Pharmacogenomics*. v. 14, p. 413-424, 2013.

WARDE-FARLEY D. et al. The GeneMania prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*. W214-220, 2010.

WEBER, J. L. et al. Human diallelic insertion/deletion polymorphisms *Am. J Hum Genet*. v.71, p. 854-862, 2002.

WEI, W. et al. Perisynaptic localization of delta subunit-containing GABA(A) receptors and their activation by GABA spillover in the mouse dentate gyrus. *J Neurosci*. v.23, p. 10650-10661, 2003.

WEINER, M. W. et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimers Dement*. v.8, p. 1-68, 2012.

YAMADA K. et al. Evidence of association between gamma-aminobutyric acid type A receptor genes located on 5q34 and female patients with mood disorders. *Neurosci Lett.* v. 349, p. 9-12, 2003.

YAN, D.; TOMITA, S. Defined criteria for auxiliary subunits of glutamate receptors. *J Physiol.* v. 590, p.21-31, 2012.

ZALSMAN, G. et al. DRD4 receptor gene exon III polymorphism in inpatient suicidal adolescents. *J Neural Transm.* v.111, p. 1593–1603, 2004.

ZHU, M.Y. et al. Elevated levels of tyrosine hydroxylase in the locus coeruleus in major depression. *Biol Psychiatry.* v.46, p. 1275–1286, 1999.

## Capítulo 2 – Artigo 1

Searching for New Genetic Variations in Expression Databases for the GABAergic and Glutamatergic Systems

# Searching for New Genetic Variations in Expression Databases for the GABAergic and Glutamatergic Systems

Manuela Barbosa Rodrigues de Souza ·  
João Ricardo Mendes de Oliveira

Received: 1 November 2011 / Accepted: 8 April 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** Changes in gene expression and genetic variations in coding regions have likely functional impact, potentially associated with complex diseases, such as neuropsychiatric conditions. A current need for high throughput analysis of genomic data is leading to the development and improvement of sophisticated bioinformatics approaches, which allows the processing of large amounts of sequence and gene expression data. In this study, we identified new potential genetic variations prioritizing genes related to glutamatergic and GABAergic systems, using different bioinformatics resources. The CLCbio Workbench Combined platform was initially used to build expressed sequence tags and mRNA files retrieved, respectively, from the Goldenpath and National Center for Biotechnology Information databases and latter to perform multiple batches of Smith–Waterman alignments. The PMUT software was used to increase an accurate association between potential variations and pathogenic predictions. The annotation revealed various classes of variations and most of them are deletions ranging from 1 to 7 bp. Bioinformatic pipelines seem to be useful approaches to help screening for genetic variations with potential impact in gene expression. Further analysis will foster this aim to provide celerity at the massive analysis of data currently generated in large scale high throughput experiments.

**Keywords** Genetic variations · Neurogenetics · Mood disorders · Bioinformatics · Microarrays · Expressed sequence tag

## Background

Multiple projects of sequencing and resequencing at the human genome reveal continuously a wide range of genetic variations such as single nucleotide polymorphism (SNP), copy number variation (CNV), deletions, insertions, inversions, and translocations (Feuk et al. 2006; Rosenfeld et al. 2010). Recently, large-scale identification and characterization of SNPs have attracted much interest, particularly in studies of complex traits like neuropsychiatric disease. A major importance is given to nonsynonymous and frameshift mutations, usually associated with likely functional impact, in addition to mutations in the untranslated region (UTR) of 5' and 3' involved in mRNA translation, RNA cleavage control, and stability (Macintyre et al. 2010; Panitz et al. 2007).

The field of bioinformatics is currently developing rapidly with the goal of assisting molecular biologists, and many web-based softwares became excellent auxiliary tools to explore genomic data mining (Teufel et al. 2006; Mooney 2005). In this article, we propose a technique to identify new genetic variations, prioritizing genes related to glutamatergic and GABAergic systems, using different bioinformatics resources. While our approach is similar to that of Lemos et al. (2009) and de Souza et al. (2010), it was added to the analysis of nonsynonymous variations at the PMUT software. Glutamate and gamma-aminobutyric acid (GABA) are, respectively, the major excitatory and inhibitory neurotransmitters in the brain, and studies show that both systems play a role at the pathophysiology of mood disorders (Schousboe and Waagepetersen 2007; Sanacora et al.

M. B. R. de Souza · J. R. M. de Oliveira  
Keizo Asami Laboratory (LIKA),  
Federal University of Pernambuco (UFPE),  
50670901 Recife, Pernambuco, Brazil

M. B. R. de Souza  
e-mail: manu.brsouza@gmail.com

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

Published online: 22 April 2012

 Springer

2008; Choudary et al. 2005; Merali et al. 2004; Krystal et al. 2002).

## Methodology

In this analysis, we extended this approach to new genes selected from microarrays studies of specific brain samples of patients with major depressive disorder (MDD) (anterior cingulate cortex) and bipolar affective disorder (BPD) (left dorsolateral prefrontal cortex). Selected genes were chosen from a reference microarray study (*GABRA5*, *GRM3*, *GRIK5*, *GABRB3*, *GABRD*, *GLUL*, and *GABRG2*), which showed altered expression in different groups of patients (Choudary et al. 2005).

The software, CLCbio Workbench Combined® version 3.6.2., was used during the following steps: building dataset, multiple batches alignment, and annotation of variations. Figure 1 summarizes the methodology employed and detailed below.

The sequences of mRNA and expressed sequence tags (ESTs) files were retrieved, respectively, from the National Center for Biotechnology Information (NCBI) ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and Goldenpath ([www.genome.ucsc.edu](http://www.genome.ucsc.edu)) databases. These sequences were used in the building of the database within the software CLCbio Workbench Combined®.

After building the database, multiple alignments have been performed, using the Smith–Waterman algorithm. This algorithm is applied for performing local sequence alignment, compares segments of all possible lengths, and optimizes the similarity measure (Smith and Waterman 1981). The Smith–Waterman algorithm is implemented by own software CLCbio Workbench Combined®. The process of multiple alignments was accelerated using dedicated processors named CUBE®.

Following the alignment, using tools already available for the software, variations were identified within the sequences of genes *GABRA5*, *GRM3*, *GRIK5*, *GABRB3*, *GABRD*, *GLUL*, and *GABRG2*. An initial number of 6,438 ESTs related to the genes previously selected were screened

to decrease the chances of handling sequencing artifact, excluding alignments with 100 % identity and e-value index under  $10^{-20}$ . A final list of 122 ESTs identity was analyzed, and after a proper annotation, we searched the NCBI SNPs database for any of the variations detected during this study.

The guanine-cytosine (GC) content, a tool available at the UCSC genome browser platform, was used to identify and avoid regions, more likely, to generate sequencing artifacts. The PMUT software (<http://mmb2.pcb.ub.es:8080/PMut/>) was used for an accurate association between potential variations and pathogenic predictions. PMUT aims to perform a fast and accurate prediction of the association of a single amino acid polymorphism (SAP) and a pathological nature. This software is based on the use of artificial neural networks, trained with a database of pathological and neutral mutations. The hit rate is 83.5 % (Ferrer-Costa et al. 2005). In addition, the PMUT also has a reliability index (ranging between 0 and 9) which can be used to keep only the best estimative (those with reliability index of >5) (Ferrer-Costa et al. 2004).

## Results

Annotation coming from our analysis predicts various classes of variations and most of them are deletions (131) ranging from 1 to 7 bp. Amongst this specific group, some were frameshift deletions (61), and the virtual translations of a few others (2) were predicted to induce an amino acid removal, with no further subsequent repercussions at the protein sequence. Possible transversions (2), transitions (17), synonymous SNPs (33), not synonymous SNPs (118), and changes in UTRs (56) were also identified (Table 1).

The transversions detected were previously described and annotated in SNP repositories (SNP db), demonstrating that our methodology was able to use EST data to detect actual DNA variations confirmed elsewhere in samples of genomic DNA, suggesting that new unreported variations can be anticipated by this method. Additionally, we noticed clusters of mismatches in different ESTs libraries in certain exons of the genes *GABRA5*, *GABRB3*, *GLUL*, and *GABRG2* (Fig. 2).

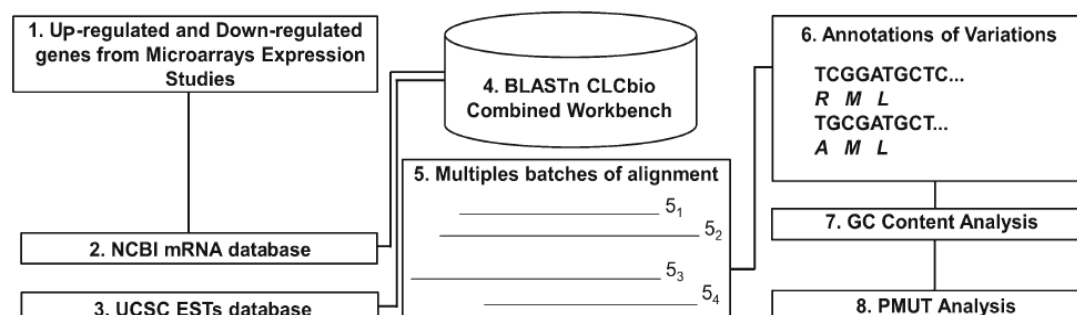


Fig. 1 Sequential steps of the proposed bioinformatics pipeline



**Table 1** Summary of in silico analysis of *GABRA5*, *GRM3*, *GRIK5*, *GABRB3*, *GABRD*, *GLUL*, and *GABRG2* genes

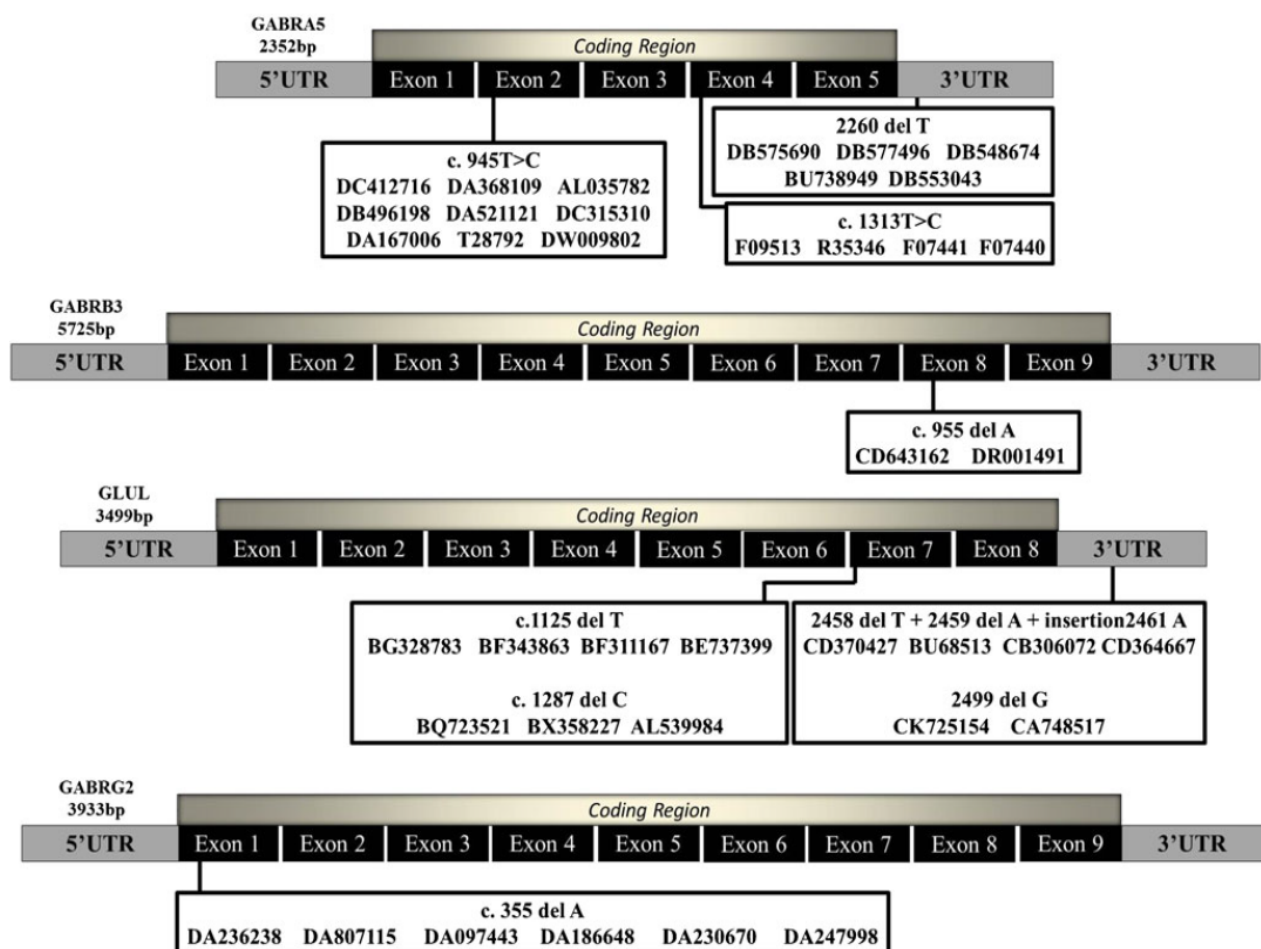
| Annotations                              | Number |
|------------------------------------------|--------|
| ESTs aligned                             | 8,482  |
| ESTs selected after stringent parameters | 137    |
| Deletions                                | 131    |
| Transversions                            | 2      |
| Transitions                              | 17     |
| Synonymous SNPs                          | 33     |
| Not synonymous SNPs                      | 118    |
| Changes in UTRs                          | 56     |

These findings make them actual more promising candidates to be explored afterwards as candidate risk factors. Two potentially novel synonymous variations at the *GABRA5* gene (exon 5/c.2304T>C) and at the *GRIK5* gene (exon 15/c.1994G>T) are not yet described. On the other hand, changes in exon 2 (c.945T>C) and exon 4 (c.1314T>C) at the

*GABRA5* gene, identified in this analysis, had already been described as rs140682 and rs140685, respectively, showing once again that this method is a feasible approach to detect genuine genetic variations that were previously confirmed in actual DNA samples.

A virtual translation detected a greater number of non-synonymous variations and some of them, caused by micro-deletions, are predicted to change the coding frame or to generate the stop codon and truncated proteins in some cases. The annotation revealed deletions ranging from 1 to 7 bp, predicted in coding regions, 5' and 3'UTR regions. The details of the annotation are shown in the Table 2.

The percentage distribution of GC content at the regions analyzed ranged below 50 % and this additional evaluation was a standard procedure during this study to select only variations in areas of low GC content, since these regions present lower chances of providing sequencing artifacts, which might be misinterpreted as true variations. The not synonymous SNPs were subjected to analysis on the website PMUT but were considered as neutral variations.



**Fig. 2** Clusters of ESTs at the positions of some predicted variations for the genes *GABRA5*, *GABRB3*, *GLUL*, and *GABRG2*. Identification of ESTs codes and positions in UTR and CDS regions are fully displayed at Table 2



**Table 2** Additional polymorphisms in *GABRA5*, *GRM3*, *GRIK5*, *GABRB3*, *GABRD*, *GLUL*, and *GABRG2* genes

| Location                                                                                                                                    | Predicted nucleotide variation<br>in coding and noncoding sequences                                                                                                                 | Predicted protein variation<br>in coding sequences                                                                                                                                                                                                                                                                              | ESTs code |
|---------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Exon 1/ <i>GABRA5</i>                                                                                                                       | c.881delC                                                                                                                                                                           | Pro180del fs                                                                                                                                                                                                                                                                                                                    | AL03572   |
|                                                                                                                                             | c.868delC                                                                                                                                                                           | 177 fs                                                                                                                                                                                                                                                                                                                          | BE8898    |
| Exon 2/ <i>GABRA5</i>                                                                                                                       | c.945T>C                                                                                                                                                                            | Synonymous                                                                                                                                                                                                                                                                                                                      | DC412716  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DA368109  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | AL035782  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DB496198  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | AL120988  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | T28792    |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DA521121  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DC315310  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DA167006  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DW009802  |
| Exon 3/ <i>GABRA5</i><br>Exon 4/ <i>GABRA5</i>                                                                                              | c.1178delT<br>c.1302delC<br>c.1313T>C                                                                                                                                               | Gly241del stop condon<br>Gly396del fs and stop introduction in position 398aa<br>Synonymous                                                                                                                                                                                                                                     | CV025335  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | R35346    |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | AL120988  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | F09513    |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | R35346    |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | F07441    |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | F07440    |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DB533989  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | AI807046  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DB575690  |
| Exon 5/ <i>GABRA5</i>                                                                                                                       | 2304T>C<br>2083delT (3'UTR)<br>2260delT (3'UTR)                                                                                                                                     | Synonymous                                                                                                                                                                                                                                                                                                                      | DB577496  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DB548674  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BU738949  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DB553043  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | DB537480  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BG706422  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BI599936  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | CD609638  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | CD609637  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BX410048  |
| Exon 2/ <i>GRM3</i><br>Exon 3/ <i>GRM3</i><br>Exon 4/ <i>GRM3</i><br>Exon 5/ <i>GRM3</i>                                                    | c.1202_1208delAATAGAA<br>c.1222delT<br>c.1830delG<br>c.2078delC<br>c.2131delA<br>c.2126delC<br>c.2955_2959delTTGTT<br>c.3017_3021delCATCT<br>c.3117delG<br>c.2477delT<br>c.3513delC | Ile35_Glu36del fs<br>Ile41del fs and stop introduction in position 41aa<br>Ala235del fs<br>Ser238del fs<br>Lys423del fs<br>Tyr334del fs<br>Leu610_Phe611del fs<br>Ile631del fs<br>Val665del fs<br>450Phe>Leu fs<br>Cys795del fs and stop introduction in position 795aa<br>Cys795del fs and stop introduction in position 795aa | CD609639  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BI753076  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BF952990  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | AL540214  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | CD609641  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BF952992  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BF952996  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BF952998  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BQ339573  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | AA670430  |
| Exon 6/ <i>GRM3</i><br>Exon 4/ <i>GRIK5</i><br>Exon 5/ <i>GRIK5</i><br>Exon 7/ <i>GRIK5</i><br>Exon 8/ <i>GRIK5</i><br>Exon 9/ <i>GRIK5</i> | c.3565delA<br>c.3526delT<br>c.3688delC<br>3814delG(3'UTR)<br>c.475_480delCTTGGC<br>c.603_605delAGT<br>c.834delC<br>c.1006delA<br>c.1186delG+1187delG                                | Lys813del fs<br>Leu800del fs<br>Pro854del fs<br><br><br>Val190del<br>Phe267del<br>Ser326del fs<br>Gly286del fs                                                                                                                                                                                                                  | DB538646  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BF305115  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BF305115  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | BF305115  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | CD300587  |
|                                                                                                                                             |                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                 | H06567    |

**Table 2** Additional polymorphisms in *GABRA5*, *GRM3*, *GRIK5*, *GABRB3*, *GABRD*, *GLUL*, and *GABRG2* genes

| Location      | Predicted nucleotide variation<br>in coding and noncoding sequences | Predicted protein variation<br>in coding sequences   | ESTs code |
|---------------|---------------------------------------------------------------------|------------------------------------------------------|-----------|
| Exon 1/GABRA5 | c.881delC                                                           | Pro180del fs                                         | AL03572   |
|               | c.868delC                                                           | 177 fs                                               | BE8898    |
| Exon 2/GABRA5 | c.945T>C                                                            | Synonymous                                           | DC412716  |
|               |                                                                     |                                                      | DA368109  |
|               |                                                                     |                                                      | AL035782  |
|               |                                                                     |                                                      | DB496198  |
|               |                                                                     |                                                      | AL120988  |
|               |                                                                     |                                                      | T28792    |
|               |                                                                     |                                                      | DA521121  |
|               |                                                                     |                                                      | DC315310  |
|               |                                                                     |                                                      | DA167006  |
|               |                                                                     |                                                      | DW009802  |
|               |                                                                     |                                                      | CV025335  |
| Exon 3/GABRA5 | c.1178delT                                                          | Gly241del stop condon                                | R35346    |
| Exon 4/GABRA5 | c.1302delC                                                          | Gly396del fs and stop introduction in position 398aa | AL120988  |
|               | c.1313T>C                                                           | Synonymous                                           | F09513    |
|               |                                                                     |                                                      | R35346    |
|               |                                                                     |                                                      | F07441    |
|               |                                                                     |                                                      | F07440    |
| Exon 5/GABRA5 | 2304T>C                                                             | Synonymous                                           | DB533989  |
|               | 2083delT (3'UTR)                                                    |                                                      | AI807046  |
|               | 2260delT (3'UTR)                                                    |                                                      | DB575690  |
|               |                                                                     |                                                      | DB577496  |
|               |                                                                     |                                                      | DB548674  |
|               |                                                                     |                                                      | BU738949  |
|               |                                                                     |                                                      | DB553043  |
|               |                                                                     |                                                      | DB537480  |
| Exon 2/GRM3   | c.1202_1208delAATAGAA                                               | Ile35_Glu36del fs                                    | BG706422  |
|               | c.1222delT                                                          | Ile41del fs and stop introduction in position 41aa   | BI599936  |
| Exon 3/GRM3   | c.1830delG                                                          | Ala235del fs                                         | CD609638  |
|               | c.2078delC                                                          | Ser238del fs                                         | CD609637  |
|               | c.2131delA                                                          | Lys423del fs                                         | BX410048  |
|               | c.2126delC                                                          | Tyr334del fs                                         | CD609639  |
| Exon 4/GRM3   | c.2955_2959delTTGTT                                                 | Leu610_Phe611del fs                                  | BI753076  |
|               | c.3017_3021delCATCT                                                 | Ile631del fs                                         | BF952990  |
|               | c.3117delG                                                          | Val665del fs                                         | AL540214  |
|               | c.2477delT                                                          | 450Phe>Leu fs                                        | CD609641  |
| Exon 5/GRM3   | c.3513delC                                                          | Cys795del fs and stop introduction in position 795aa | BF952992  |
|               |                                                                     | Cys795del fs and stop introduction in position 795aa | BF952996  |
|               | c.3565delA                                                          | Lys813del fs                                         | BF952998  |
|               | c.3526delT                                                          | Leu800del fs                                         | BQ339573  |
| Exon 6/GRM3   | c.3688delC                                                          | Pro854del fs                                         | AA670430  |
|               | 3814delG(3'UTR)                                                     |                                                      | DB538646  |
| Exon 4/GRIK5  | c.475_480delCTTGGC                                                  |                                                      | BF305115  |
| Exon 5/GRIK5  | c.603_605delAGT                                                     | Val190del                                            | BF305115  |
| Exon 7/GRIK5  | c.834delC                                                           | Phe267del                                            | BF305115  |
| Exon 8/GRIK5  | c.1006delA                                                          | Ser326del fs                                         | CD300587  |
| Exon 9/GRIK5  | c.1186delG+1187delG                                                 | Gly286del fs                                         | H06567    |

**Table 2** (continued)

| Location      | Predicted nucleotide variation<br>in coding and noncoding sequences | Predicted protein variation<br>in coding sequences | ESTs code |
|---------------|---------------------------------------------------------------------|----------------------------------------------------|-----------|
| Exon 10/GRIK5 | c.1268_1271delCACT                                                  | Thr3314del fs                                      | H06567    |
| Exon 11/GRIK5 | c.1365delC+1366delG                                                 |                                                    | AL048958  |
| Exon 13/GRIK5 | c.1629delG                                                          |                                                    | AL048958  |
| Exon 14/GRIK5 | c.1777delG                                                          | Ala486del fs                                       | AA977136  |
| Exon 15/GRIK5 | c.1944G>T                                                           |                                                    | BM551231  |
|               | c.1957delG                                                          | Ala549del fs                                       | T64839    |
|               | c.1957delG                                                          | Ala549del+Phe550del fs                             | R60372    |
| Exon 16/GRIK5 | c.2168_2172delGCGTC                                                 | Arg620del+Val621del fs                             | R59240    |
| Exon 19/GRIK5 | c.2646delG                                                          | Ser693del fs                                       | BI964037  |
|               | 3274_3279delGAGGAC(3'UTR)                                           |                                                    | AA971174  |
| Exon 5/GABRB3 | c.339delT+340delA                                                   | Asp94>Val fs                                       | DN992028  |
|               | c.412delG                                                           | Gln115>His fs                                      | F05759    |
| Exon 7/GABRB3 | c.631delT                                                           | Tyr194del fs                                       | BE780160  |
| Exon 8/GABRB3 | c.860delA+861delT                                                   | Asn269del fs                                       | CD652458  |
| Exon 9/GABRB3 | c.955delA                                                           | Ile300del fs                                       | CD643162  |
|               |                                                                     |                                                    | DR001491  |
| Exon 1/GABRD  | 95delC(5'UTR)                                                       |                                                    | BI824761  |
| Exon 2/GABRD  | c.165delG                                                           | Glu37del fs                                        | DA071101  |
|               | c.213delC                                                           | Arg53del fs                                        | W81526    |
| Exon 6/GABRD  | c.771delG                                                           | Met226del fs                                       | BI757337  |
|               | c.760delA                                                           | Thr222del fs                                       | BM545512  |
|               | c.695delA                                                           | Gln200del fs                                       | BG037021  |
|               | c.744delC                                                           | Thr217del fs                                       | AV122311  |
| Exon 7/GABRD  | c.885delC                                                           | Met264del fs                                       | BI829118  |
|               | c.817_819delTTC                                                     | Phe241del fs                                       | H41122    |
|               | c.829delA                                                           | Arg245del fs                                       | BI758571  |
| Exon 1/GABRD  | 95delC(5'UTR)                                                       |                                                    | BI824761  |
| Exon 8/GABRD  | c.1089delC                                                          | Ala331del fs                                       | H41122    |
| Exon 9/GABRD  | c.1293delG                                                          | Glu600del fs                                       | BI822807  |
| Exon 8/GLUL   | c.1125delT                                                          | 205Ile>Met fs                                      | BG328783  |
|               |                                                                     |                                                    | BF343863  |
|               |                                                                     |                                                    | BF311167  |
|               |                                                                     |                                                    | BE737399  |
|               | c.1135delG                                                          | Glu210del fs                                       | BQ339490  |
|               | c.1127_1129delGAC                                                   | Gly206del fs                                       | AV758685  |
|               | c.1171delC                                                          | Arg222del fs                                       | AV758682  |
|               | c.1203delT                                                          | 231Phe>Leu fs                                      | BG286627  |
|               | c.1218delC                                                          | Phe298del fs                                       | BX395398  |
|               | c.1259delG                                                          | Gly251del fs                                       | AL548328  |
|               | c.1279delA                                                          | Thr258del fs                                       | AL570809  |
|               | c.1268_1212delCAAC                                                  | Thr254del+Asn255del fs                             | AL553426  |
|               | c.1264delC                                                          | His253del fs                                       | BI869492  |
|               | c.1275delC+1276delA                                                 | 256Phe>Leu fs                                      | BU526529  |
|               | c.1283>Ap/G                                                         | 259Lys>Arg fs                                      | BM473826  |
|               | c.1287delC                                                          | Met261del fs                                       | BQ723521  |
|               |                                                                     |                                                    | BX358227  |
|               |                                                                     |                                                    | AL539984  |
| Exon 9/GLUL   | c.1476delA                                                          | Arg354del fs                                       | BX411411  |

Table 2 (continued)

| Location      | Predicted nucleotide variation<br>in coding and noncoding sequences | Predicted protein variation<br>in coding sequences | ESTs code |
|---------------|---------------------------------------------------------------------|----------------------------------------------------|-----------|
|               | 2089delG(3'UTR)                                                     |                                                    | AL576337  |
|               | 2956delA(3'UTR)                                                     |                                                    | CF455674  |
|               | 2136delG(3'UTR)                                                     |                                                    | AL574476  |
|               | 2239delG(3'UTR)                                                     |                                                    | BM476840  |
|               | 2593delT(3'UTR)                                                     |                                                    | BU164034  |
|               | 2267delA+2268delC(3'UTR)                                            |                                                    | BQ670907  |
|               | 2041delT(3'UTR)                                                     |                                                    | BU165515  |
|               | 2458Tdel+2459A+insertion 2461A(3'UTR)                               |                                                    | CD370427  |
|               |                                                                     |                                                    | BU685135  |
|               |                                                                     |                                                    | CB306072  |
|               |                                                                     |                                                    | CD364667  |
|               | 1959delT(3'UTR)                                                     |                                                    | BG403248  |
|               | 3804delA(3'UTR)                                                     |                                                    | BG290718  |
|               | 2037_2040delGTGG (3'UTR)                                            |                                                    | CD517441  |
|               | 2328delT(3'UTR)                                                     |                                                    | BI827221  |
|               | 2365delA+2366delT(3'UTR)                                            |                                                    | CF457060  |
|               | 2891delT(3'UTR)                                                     |                                                    | BG567715  |
|               | 2898delT(3'UTR)                                                     |                                                    | BU181118  |
|               | 1372delG(3'UTR)                                                     |                                                    | AL550535  |
|               | 2480delT(3'UTR)                                                     |                                                    | AW074130  |
|               | 2499delG(3'UTR)                                                     |                                                    | CK725154  |
|               |                                                                     |                                                    | CA748517  |
|               | 3108delG+3109delT(3'UTR)                                            |                                                    | BQ717088  |
|               | 1901delC(3'UTR)                                                     |                                                    | BF339029  |
|               | 2104delG(3'UTR)                                                     |                                                    | BE782083  |
|               | 2191delG(3'UTR)                                                     |                                                    | BQ879479  |
| Exon 1/GABRG2 | c.355delA                                                           | Ala122del fs                                       | DA236238  |
|               |                                                                     |                                                    | DA807115  |
|               |                                                                     |                                                    | DA097443  |
|               |                                                                     |                                                    | DA186648  |
|               |                                                                     |                                                    | DA230670  |
|               |                                                                     |                                                    | DA247998  |
|               | c.353_355delAAG                                                     |                                                    | DA807093  |
| Exon 3/GABRG2 | c.661delT                                                           | 222Ile>Met fs                                      | BI548984  |
|               | c.630A>T                                                            | 212 stop codon                                     | DB466433  |
| Exon 4/GABRG2 | c.736delG                                                           | Lys248del fs                                       | DA 416076 |
|               | c.749delA                                                           | Thr252del fs                                       | DA495517  |
|               | c.842delT                                                           | Thr283del fs                                       | DA093933  |
|               | c.857_860delAACA                                                    | Asn288del fs                                       | BI600201  |
| Exon 7/GABRG2 | c.929delT                                                           | 310>Ile Pro Met                                    | BI544530  |
|               | c.951delT                                                           | 317Serdel fs                                       | BI597864  |

fs frameshift

## Discussion

More recently, several studies have associated changes in glutamatergic and GABAergic systems of neurotransmitters

in the pathophysiology of mood disorders (Sanacora et al. 2008; Karolewicz et al. 2010; Shen et al. 2010). Chronic antidepressant use and electroconvulsive therapy regulates ionotropic *N*-methyl-D-aspartate (NMDA) receptors



expression and function. Valproic acid, a mood stabilizer frequently prescribed for bipolar disorder (manic depressive illness), has different effects on the glutamatergic system, regulating glutamate and GABA transporter protein, changing also the vesicular glutamate transporter expression and presynaptic glutamate release, and modulating glutamate receptor expression and function (Sanacora et al. 2008).

On the other hand, the anxiety symptoms present in most depressive syndromes are often treated with benzodiazepines agents, such as diazepam and clonazepam, acting at the GABAergic system. However, another hypothesis has considered the GABAergic deficit to play a central role at the pathogenesis of major depression (Rudolph and Knoflach 2011; Luscher et al. 2011).

The genes analyzed in this study were previously investigated with different approaches and have been involved with the glutamatergic and GABAergic systems in mood disorders (Choudary et al. 2005). Genetic variations might contribute to differences in protein activities and differential gene expression in the brain of individuals with mood disorders. Single nucleotide polymorphisms are the most common type of genetic variations in humans and are thought to be the genetic basis or markers for most human diseases (de Souza et al. 2010; Choi et al. 2011; Capriotti et al. 2006; Reumers et al. 2009).

Recent studies show level changes altered in *GLUL* gene in the prefrontal cortex of suicidal MDD subjects (Sequeira et al. 2009). Performing a comparison of gene expression profiles of postmortem brains of patients with mental illness, there was a tendency in a decreased expression of genes that encode ionotropic glutamate (Iwamoto et al. 2004).

Variations in genes *GABRD* and *GABRA5* have been associated with several psychiatric disorders, as noted in a recent study where variations in both genes show strong evidence of risk in BPD (Craddock et al. 2010). Mutations in the gene *GABRB3* have been associated with neuropsychiatric disorders, such as autism, MDD, BPD, and schizophrenia (Buxbaum et al. 2002; DeLong 2004).

Interestingly, the rs6465084 SNP at the *GRM3* gene was recently detected in Japanese patients with MDD (Tsunoka et al. 2009). Previous studies have related various indels as risk factors for mood disorders. The 5-HTTLPR polymorphism, for example, is a 44-bp indel within the promoter region of the serotonin transporter gene, and the short variant is considered the risk factor for different mood disorders and anxiety traits. A 19-bp insertion/deletion polymorphism in the dopamine beta-hydroxylase (DBH) gene has been also studied in mood disorders and drug addiction. Our work might unveil new indels with potential impact in helping to explain how important are such variations for the clinical outcome of affective conditions (Togsvær et al. 2008; Oliveira et al. 2000).

Bioinformatic pipelines seem to be useful approaches to help screening for genetic variations with potential impact in gene expression; however, there is a need to optimize this use in more integrated platforms, preferentially with the full processing in one single virtual environment which is not available yet. Curiously, the pattern of variation found in this study was similar to the pattern of variation found in previous studies that used the software CLCbio and hardware CUBE in search of new variations in candidate genes for Alzheimer's disease, major depressive disorder, bipolar affective disorder, and sporadic Creutzfeldt–Jakob disease (Lemos et al. 2009; de Souza et al. 2010), showing once again that this method is a feasible approach to detect genetic variations.

The use of different softwares to predict the pathogenicity of genetic variations is always debatable because of the often discrepancy between some of them, even when analyzing the same variation. We recently applied such analysis to better comprehend a missense substitution candidate for Fahr's disease at the *MGEA6* gene, in a proline-rich and highly conserved protein domain. SAP, PolyPhen, and I-mutant suggest that this variation is predicted to cause probable molecular deleterious effects, but the PMUT considered this same mutation as neutral. The SIFT program generated conflicting findings, depending if we used the UniProt-SWISS PROT 56.6, NCBI, or UniProt-TrEMBL 39.6 databases. The discrepancies might be due to the limited skill of these programs in the prediction or the poor source of data available for structural model to the proteic product (Lemos et al. 2011).

This is why it is important to always focus also in methods that try to confirm the predicted variation and latter validate the predicted pathogenicity. Further analysis will foster this aim to provide celerity at the massive analysis of data currently generated in large-scale high throughput experiments.

## Conclusion

After using a sequence of steps associated with the various tools of bioinformatics, we observed that the study of EST sequences to identify new variations of candidates is a promising approach and accessible. We have found variations previously described, including probable candidate polymorphisms previously unpublished, for the manifestation of mood disorders.

Therefore, these findings raise the possibility that deletions of up to 7 bp are related to the pathophysiology of TH, possibly in genes linked to GABAergic and glutamatergic systems of neurotransmitters.

**Acknowledgments** We are greatly indebted to Henrique Castelletti and Roberta Rodrigues de Lemos for their technical support. This

study received financial support from the following Brazilian funding agencies and academic bureaus: LIKA-JIKA, PROPESQ-UFPE, CAPES, CNPq, and FACEPE.

**Conflict of interest** There is no conflict of interest in this study.

## References

- Buxbaum JD, Silverman JM, Smith CJ et al (2002) Association between a GABRB3 polymorphism and autism. *Mol Psychiatry* 7:311–316
- Capriotti E, Calabrese R, Casadio R (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* 22:2729–2734
- Choi KH, Higgs BW, Wendland JR et al (2011) Gene expression and genetic variation data implicate PCLO in bipolar disorder. *Biol Psychiatry* 69(4):353–359
- Choudary PV, Molnar M, Evans SJ et al (2005) Altered cortical glutamatergic and GABAergic signal transmission with glial involvement in depression. *Proc Natl Acad Sci U S A* 102:15653–15658
- Craddock N, Jones L, Jones IR et al (2010) Strong genetic evidence for a selective influence of GABAA receptors on a component of the bipolar disorder phenotype. *Mol Psychiatry* 15:146–153
- de Souza MB, de Lemos RR, da Cunha JE, de Lima Filho JL, de Oliveira JR (2010) Searching for new genetic risk factors for neuropsychiatric disorders in expression databases. *J Mol Neurosci* 41:193–197
- DeLong R (2004) Autism and familial major mood disorder: are they related? *J Neuropsychiatry Clin Neurosci* 16:199–213
- Ferrer-Costa C, Orozco M, de la Cruz X (2004) Sequence-based prediction of pathological mutations. *Proteins* 57:811–819
- Ferrer-Costa C, Orozco M, de la Cruz X (2005) Use of bioinformatics tools for the annotation of disease-associated mutations in animal models. *Proteins* 61:878–887
- Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7:85–97
- Iwamoto K, Kakiuchi C, Bundo M, Ikeda K, Kato T (2004) Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. *Mol Psychiatry* 9:406–416
- Karolewicz B, Maciag D, O'Dwyer G, Stockmeier CA, Feyissa AM, Rajkowska G (2010) Reduced level of glutamic acid decarboxylase-67 kDa in the prefrontal cortex in major depression. *Int J Neuropsychopharmacol* 13:411–420
- Krystal JH, Sanacora G, Blumberg H et al (2002) Glutamate and GABA systems as targets for novel antidepressant and mood-stabilizing treatments. *Mol Psychiatry* 7(Suppl 1):S71–S80
- Lemos R, Castelletti CH, Lima Filho JL, Marques E, Oliveira JR (2009) In silico identification of new genetic variations as potential risk factors for Alzheimer's disease in a microarray-oriented simulation. *J Mol Neurosci* 39:241–247
- Lemos R, Oliveira DF, Zatz M, Oliveira JR (2011) Population and computational analysis of the MGEA6 P521A variation as a risk factor for familial idiopathic basal ganglia calcification (Fahr's disease). *J Mol Neurosci* 43:333–336
- Luscher B, Shen Q, Sahir N (2011) The GABAergic deficit hypothesis of major depressive disorder. *Mol Psychiatry* 16:383–406
- Macintyre G, Bailey J, Haviv I, Kowalczyk A (2010) is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics* 26:524–530
- Merali Z, Du L, Hrdina P et al (2004) Dysregulation in the suicide brain: mRNA expression of corticotropin-releasing hormone receptors and GABA(A) receptor subunits in frontal cortical brain region. *J Neurosci* 24:1478–1485
- Mooney S (2005) Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. *Brief Bioinform* 6:44–56
- Oliveira JR, Carvalho DR, Pontual D et al (2000) Analysis of the serotonin transporter polymorphism (5-HTTLPR) in Brazilian patients affected by dysthymia, major depression and bipolar disorder. *Mol Psychiatry* 4:348–349
- Panitz F, Stengaard H, Hornshøj H et al (2007) SNP mining porcine ESTs with MAVIANT, a novel tool for SNP evaluation and annotation. *Bioinformatics* 23:387–391
- Reumers J, Schymkowitz J, Rousseau F (2009) Using structural bioinformatics to investigate the impact of non synonymous SNPs and disease mutations: scope and limitations. *BMC Bioinform* 10(Suppl 8):S9
- Rosenfeld JA, Malhotra AK, Lencz T (2010) Novel multi-nucleotide polymorphisms in the human genome characterized by whole genome and exome sequencing. *Nucleic Acids Res* 38:6102–6111
- Rudolph U, Knoflach F (2011) Beyond classical benzodiazepines: novel therapeutic potential of GABAA receptor subtypes. *Nat Rev Drug Discov* 10:1–13
- Sanacora G, Zarate CA, Krystal JH, Manji HK (2008) Targeting the glutamatergic system to develop novel, improved therapeutics for mood disorders. *Nat Rev Drug Discov* 7:426–437
- Schousboe A, Waagepetersen HS (2007) GABA: homeostatic and pharmacological aspects. *Prog Brain Res* 160:9–19
- Sequeira A, Mamdani F, Ernst C et al (2009) Global brain gene expression analysis links glutamatergic and GABAergic alterations to suicide and major depression. *PLoS One* 4:65–85
- Shen Q, Lal R, Luellen BA, Earnheart JC, Andrews AM, Luscher B (2010) gamma-Aminobutyric acid-type A receptor deficits cause hypothalamic-pituitary-adrenal axis hyperactivity and antidepressant drug sensitivity reminiscent of melancholic forms of depression. *Biol Psychiatry* 68:512–522
- Smith TF, Waterman MS (1981) Identification of common molecular subsequences. *J Mol Biol* 147:195–197
- Teufel A, Krupp M, Weinmann A, Galle PR (2006) Current bioinformatics tools in genomic biomedical research (Review). *Int J Mol Med* 17:967–973
- Togsværød M, Werge T, Tankó LB et al (2008) Association of a dopamine beta-hydroxylase gene variant with depression in elderly women possibly reflecting noradrenergic dysfunction. *J Affect Disord* 106:169–172
- Tsunoka T, Kishi T, Ikeda M et al (2009) Association analysis of group II metabotropic glutamate receptor genes (GRM2 and GRM3) with mood disorders and fluvoxamine response in a Japanese population. *Prog Neuropsychopharmacol Biol Psychiatry* 33:875–879

## Apêndice do Capítulo 2

Troubleshooting the search for potential pathogenic small *INDELs* (1-10bp) in neuropsychiatric disorders



## Troubleshooting the search for potential pathogenic small INDELs (1-10bp) in neuropsychiatric disorders

**Lemos, RR<sup>1</sup>; Souza, MBR<sup>1</sup>; Oliveira, DF<sup>1</sup>; Pimentel, LF<sup>1</sup>; Oliveira, JRM<sup>1,2</sup>.**

<sup>1</sup>Keizo Asami Laboratory – LIKA – Federal University of Pernambuco, Recife-PE. <sup>2</sup>Neuropsychiatry Department Federal University of Pernambuco, Recife-PE.

robertalemos775@hotmail.com

**Keywords:** INDELs, Sequencing, Neuropsychiatric disorder

Single Nucleotide Polymorphisms (SNPs) and Insertion or Deletions (INDELs) have been increasingly recognized as a major type of genetic variations, with potential impact in protein activities and gene expression levels observed in complex genetic traits, like neuropsychiatric diseases. However, previous studies have pointed various methodological difficulties to reliably detect and validate short INDELs in a large-scale manner. Various approaches have been developed with bioinformatics tools and molecular techniques for identification and association of these variations and our group recently designed a bioinformatics pipeline to detect new potential variations in genes selected from microarrays studies performed with brain samples from patients with Alzheimer's disease (AD) and mood disorders (dysthymia and bipolar affective disorder-BPD). We started screening this panel of variations in Brazilian population samples affected by AD and mood disorders to confirm the variations predicted by our original method. The DNA was amplified by PCR and sequenced in a Mega Bace 1000 and fragments were aligned by using the software CLCbio Workbench Combined® version 3.6.2. So far, INDELs predicted *In silico*, ranging from 1 to 10bp, were not confirmed in any samples. Some false positives were consistently identified but latter ruled out after verification at the complementary strand. We performed preliminary experiments with DNA pooling and they still need methodological adjustments for a better interpretation for the yield and sensitivity but are potentially applicable to save time and resources in such large scale screening. Two genetic variations at the genes COX2 and IL1A were confirmed in AD samples: respectively rs2745557 and rs17561. The rs14068 SNP, located at exon 2 of the GABRA5 gene was validated in samples of patients with dysthymia and BPD. Our group is currently dealing with different approaches that might optimize sequencing and bioinformatics analyses of shorts INDELs to broaden our research capabilities of identifying those intriguing genetic variations. Reporting these pitfalls might contribute with other groups that are planning to search for such interesting genetic variation and call the attention for the high rate of false positives in case of experiments lacking careful checking of double strand sequencing. Financial Support: CNPq, PROPESQ-UFPE, CAPES, and FACEPE.



## Capítulo 3 – Artigo 2

### Exploring the Implications of *INDELs* in Neuropsychiatric Genetics: Challenges and Perspectives

## Exploring the Implications of INDELs in Neuropsychiatric Genetics: Challenges and Perspectives

R. R. Lemos · M. B. R. Souza · J. R. M. Oliveira

Received: 5 December 2011 / Accepted: 24 January 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** The decade passed after publishing the Human Genome first draft faced an enormous growth at the understanding of the genomic variation among different subjects, populations, and groups of patients. Single nucleotide polymorphisms (SNPs) and insertion or deletions (INDELs) have been increasingly recognized as a major type of genetic variations, with potential impact in protein activities and gene expression changes observed in complex genetic traits, like neuropsychiatric diseases. INDELs represent the second most common class of variations after SNPs, but there is still an important gap between the number of INDELs reported and the actual knowledge about the functional implications of such variations. There are approximately 10 million SNPs already reported, and the human populations are expected to collectively harbor at least 1.6–2.5 million INDELs. One of the major challenges is to find better platforms to screen for INDELs in a high throughput manner. The discordance in between the data from different studies might be explained by the diverse approaches employed to sequence the genomes with variable platforms. Short INDEL variations increased the scope of genetic markers in human genetic diseases, and various studies showed that common micro-deletions and smaller INDELs might be highly associated with neuropsychiatric diseases such as schizophrenia, autism, mental retardation, and Alzheimer disease. The rapidly increasing amount of resequencing, genotyping, and personal

genome data generated by large-scale genetic human projects require the development of integrated bioinformatics tools able to efficiently manage and analyze these genetic data. Our group is currently dealing with different approaches that might optimize sequencing and bioinformatics analyses of short INDELs to broaden our research capabilities of identifying those intriguing genetic variations. Hopefully, INDELs might become a new trend in association studies in neuropsychiatric genetics since so far the level of significant and positive associations with the standard SNPs reported presents limited predictive application.

**Keywords** INDELs · Bioinformatics · Neuropsychiatric disorder · Variations · Sequencing

### Introduction

The decade passed after releasing the Human Genome first data set faced an enormous growth at the understanding of the genomic variation among different subjects, populations, and groups of patients. After Lander et al. (2001) published the first complete genome human draft, a promising season started in searching for the biological functions related to the genetic code. Since that time, sequencing platforms with high-throughput technologies are generating massive amounts of genetic data, including comparative genetic variation detection among different subjects and ethnical groups. These projects are often referred as “resequencing” because they review, validate, and enrich the original genome sequence initially released (Snyder et al. 2010).

Sequencing projects generally categorize classes of variations as single nucleotide polymorphisms (SNPs), insertion/deletion (INDELs), and large structural variation, although

R. R. Lemos · M. B. R. Souza · J. R. M. Oliveira  
Keizo Asami Laboratory (LIKA),  
Federal University of Pernambuco,  
50670901 Recife, Pernambuco, Brazil

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

other nomenclatures might be used, and some authors might disagree in terms of definition (Mullaney et al. 2010). SNPs and INDELs may be used as markers in the search for genetic factors that influence neuropsychiatric diseases but also as resources for population genetics and evolutionary studies and are used routinely as markers in agricultural breeding programs (DePristo et al. 2011; Gupta et al. 2001; Sawyer et al. 2003). The limited rate of mutation in the human genome also makes these variations suitable as excellent polymorphic markers to study complex diseases (Sylvänen 2001).

There are 3–4 million SNPs present in a given subject, distributed across the genome (Ahn et al. 2009; Levy et al. 2007; Pushkarev et al. 2009; Wang et al. 2008; Wheeler et al. 2008). The approaches with massively parallel next generation sequencing have progressed rapidly over the past few years, and short INDEL variations increased the scope of genetic markers in human genetic disease (Ku et al. 2010).

Insertions or deletions represent the second most common class of variations after SNPs, but there is still an important gap between the number of INDELs reported and the functional implications of such variations. There are approximately 0.3–0.6 million INDELs ranging from 2 up to 1,000 bp in a given subject, when compared to a reference genome, but this number decreases dramatically to 320 for INDELs between 1 and 2 kb. Diallelic INDELs are abundant, accounting for approximately 8% of all human DNA sequence variants (Levy et al. 2007; Weber et al. 2002).

Small INDELs also have been found as the second frequent type mutation in personal human genomes. What was observed is that there is a divergence in both the number and size of INDELs reported in these genomes (Table 1). Curiously, 823,396 INDELs were discovered in the Craig Venter's genome, while 222,718 INDELs were found in the James Watson's genome, and 135,262 INDELs were discovered in the Han Chinese genome. Some differences were observed with the Venter's INDELs ranging from 1 to 82,711 bp in length, whereas those discovered in Han Chinese genome were limited to 1–3 bp in length. The discordance in between the data might be explained by the different approaches employed to sequence the genomes. While the Venter's genomes were sequenced through traditional application binary interface (ABI) reads, the analysis of the Han Chinese

genome employed Illumina technology (Ahn et al. 2009; Kim et al. 2009; Levy et al. 2007; Schuster et al. 2010; Wang et al. 2008; Wheeler et al. 2008).

In 2008, Kidd et al. (Kidd et al. 2008) compared the whole genome of eight subjects and found 1,695 polymorphic sites, with a total of 747 deletions, 724 insertions, and 224 inversions. Ng et al. (Ng et al. 2010) used those sequence files to compare the genomes of four unrelated individuals affected by Freeman–Sheldon syndrome (OMIM #193700). This ascertainment yielded on average 166 coding INDELs called per subject, with 63% previously annotated in dbSNP and a higher presence of INDEL multiples of 3 (“3n”) (Mills et al. 2006).

The initial map of human INDEL variations ranged between 1 and 9,989 bp in length and was described with a computational approach using DNA resequencing traces that were originally generated for SNP discovery projects, and there was an important variability when comparing the different approaches and different populations such as 374,355 INDELs using whole genome shotgun traces (HapMap consortium DNA of 8 African-Americans), 137,526 from the SNP consortium (24 subjects), and 17,217 whole chromosome 20 (shotgun WCS traces of 4 subjects) (Mills et al. 2006). Some consortiums such as the International SNP Map Working Group (2001), the International HapMap consortium (2003, 2005) and the 1000 Genomes Project consortium (2010) have elucidated small INDELs ranging from 1 bp to 1 kb in length.

The SNP database contains 10 million SNPs deposited, and different human populations are expected to collectively harbor at least 1.6–2.5 million INDEL polymorphisms ([www.ncbi.nlm.nih.gov/SNP](http://www.ncbi.nlm.nih.gov/SNP), build 125). A recent study report lists almost 2 million INDELs ranging from 1 to 10,000 bp found at the genomes of 79 diverse humans, including 819,363 small INDELs that map to human genes (Mills et al. 2011).

## INDELs in Neuropsychiatric Diseases

Previous studies have related INDELs as risk factors for neuropsychiatric diseases, such as mood disorders. Table 2

**Table 1** Divergence in both the number and size of INDELs reported in personal genome

| Study                  | Individual   | Number of INDELs | INDEL size range (bp) | Method           |
|------------------------|--------------|------------------|-----------------------|------------------|
| (Levy et al. 2007)     | Venter       | 823,396          | 1–82,711              | ABI              |
| (Wang et al. 2008)     | Han Chinese  | 135,262          | 1–3                   | Illumina/SOAP    |
| (Wheeler et al. 2008)  | Watson       | 222,718          | 2–38,896              | 454              |
| (Ahn et al. 2009)      | Korean (SJK) | 342,965          | 1–26                  | Illumina/MAQ     |
| (Kim et al. 2009)      | Korean (AK1) | 170,202          | 1–29                  | Illumina/Alpheus |
| (Schuster et al. 2010) | African      | Not reported     | Not reported          | Not reported     |



**Table 2** INDELs in relevant genes associated neuropsychiatric diseases

| Size/Class                                                                                                           | Disease                                  | Reference                                             |
|----------------------------------------------------------------------------------------------------------------------|------------------------------------------|-------------------------------------------------------|
| 44-bp INDEL promoter region of the serotonin transporter gene 5HTTLPR                                                | Mood disorders and anxiety               | (Armbruster et al. 2009)                              |
| 19-bp INDEL in the dopamine beta-hydroxylase (DBH)                                                                   | Mood disorders and drug addiction        | (Togsverd et al. 2008)                                |
| c.795delC in OGG1 gene                                                                                               | Alzheimer disease                        | (Mao et al. 2007)                                     |
| Deletions ranging from 1 to 7 bp ASPM and CENPJ genes                                                                | Autosomal recessive primary microcephaly | (Bond et al. 2003; Gul et al. 2006; Leal et al. 2003) |
| Insertion of octapeptide repeats (OPR), encompassing the codon residues 51 through 91 in the prion protein gene PRNP | Genetic Creutzfeldt–Jakob disease (gCJD) | (Capellari et al. 2011)                               |
| Deletions ranging in size from 5 to >260 kb                                                                          | Autism                                   | (Yu et al. 2002)                                      |
| Microdeletions and microduplications >100 kb                                                                         | Schizophrenia                            | (Walsh et al. 2008)                                   |

shows some INDELs in relevant genes that associate neuropsychiatric diseases. The 5-HTTLPR polymorphism, for example, is a 44-bp INDEL within the promoter region of the serotonin transporter gene, and the short variant is considered a risk factor for different mood disorders and anxiety (Armbruster et al. 2009). Another example is the 19-bp insertion/deletion polymorphism in the dopamine beta-hydroxylase gene associated with mood disorders and drug addiction (Togsverd et al. 2008).

Our group studied the 5-HTTLPR INDEL in various neuropsychiatric conditions such as schizophrenia, bipolar disorder, dysthymia, major depression, and Alzheimer's disease (Oliveira et al. 2000; Oliveira et al. 1998). However, the major current challenge is to unveil new INDELs with potential impact in helping to explain how important are such variations for the clinical outcome of behavioral and neurological conditions. This is particularly crucial because the current associated INDELs present a high rate of inconsistent associations.

Various studies showed that common microdeletions might be highly associated with neuropsychiatric diseases such as schizophrenia, autism, and mental retardation (Walsh et al. 2008; Yu et al. 2002). Interestingly, a variation c.795delC in OGG1 gene was detected in Alzheimer's disease (AD) patients, changing the coding frame and adding 59 amino acid residues downstream after the termination codon, causing reduced enzymatic activity (Mao et al. 2007).

Truncated protein products caused by deletions ranging from 1 to 7 bp were also observed in abnormal spindle-like, microcephaly-associated (ASPM) and centromere protein J (CENPJ) genes in autosomal recessive primary microcephaly (Bond et al. 2003; Gul et al. 2006; Leal et al. 2003). Other mutations between 1 and 7 bp at the ASPM gene were reported in cohorts of children with microcephaly, and so far, there are 57 variations, including 28 deletions or insertions (Nicholas et al. 2009). The pathogenic mutations linked to genetic Creutzfeldt–Jakob disease and fatal insomnia revealed that 10–15% of subjects who develop a prion

disease carry either a point mutation or an insertion of octapeptide repeats, encompassing the codon residues 51 through 91 in the prion protein gene PRNP (Capellari et al. 2011).

### Methods Detecting INDELs

To detect and assess the potential biological functions of the large number of INDELs in the human genome, it is essential to use specific and effective methodology especially the use of computational methods that facilitate and streamline the implementation process of analyzing these polymorphisms, with a cooperative integration of different genomic technologies to improve the precision and detection rate of unusual polymorphisms (Hong et al. 2011; Kim et al. 2009). Thus, several new methods have been proposed for detecting SNPs and INDELs in human genome; however, polymerase chain reaction (PCR) followed by direct automatic sequencing of genomic DNA is still the main approach that has proven effective for the identification of these polymorphisms (Bhangale et al. 2005; Bhangale et al. 2006; Mullaney et al. 2010).

The next-generation DNA sequencing technologies are powerful alternatives, and this technology promotes the DNA sequencing analysis platforms capable of generating information on million of base pairs in a single run. Among the new sequencing platforms still used for INDEL detection are 454 FLX/Roche and Illumina platforms (Ahn et al. 2009; Bentley et al. 2008; Ley et al. 2008; Wang et al. 2008; Wheeler et al. 2008). Next-generation sequencing is currently the most effective methodology used to detect and characterize INDELs; however, additional studies show that computational approaches used for next-generation trace mapping and INDEL detection have high levels of sensitivity but have a considerably high rate of false negative (Mullaney et al. 2010). Programs used for trace mapping and variation discovery in Illumina-based personal genome projects, such as SOAP, ELAND (Illumina), and MAQ,

presented lower rates of false negative, between 10% and 35% (Li et al. 2008a, b). Thus, up to a third of the small INDELs can be missed with these approaches.

Another method used is tagged microarray marker, allowing high-throughput differentiation between predicted alternative PCR products. Typically, it is used as a molecular marker approach to determine the allelic states of INDEL alleles at genomic loci in multiple individuals (Jing et al. 2007). Established methods such as ABI trace mapping and BAC overlap also show efficiency in detecting INDELs (Kidd et al. 2008; Levy et al. 2007; Mills et al. 2006; Mullikin et al. 2000).

Previous studies have pointed various methodological difficulties to reliably detect and validate short INDELs in a large-scale manner, and various approaches have been developed with bioinformatic tools and molecular techniques for identification and association of these variations. The main challenge right now is to effectively confirm the INDELs predicted by such methods and handling the false positives and false negatives in order to decrease error rate (Salathia et al. 2007). DNA pooling is also a promising alternative to overcome such limitations and simultaneous processing of multiple samples (McKernan et al. 2009).

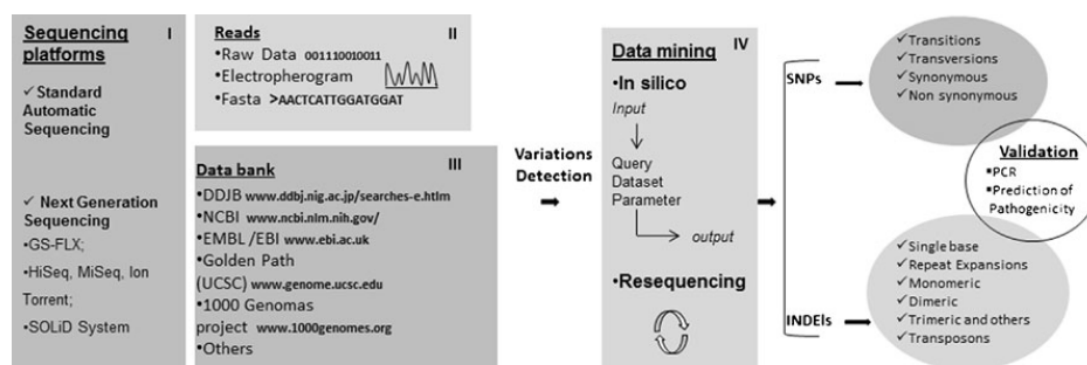
### INDELs and Computational Tools

The rapidly increasing amount of resequencing, genotyping, and personal genomes generated by large-scale genetic human projects requires the development of integrated bioinformatic tools which are able to efficiently manage and analyze such overwhelming plethora of data. Our group recently designed a bioinformatic pipeline to detect new potential variations in genes selected from studies of expression microarrays. The *in silico* analysis revealed deletions ranging from 1 to 10 bp, distributed along coding regions, 5' and 3', and UTR regions, in genes selected from microarray

studies of brain samples of patients with AD, major depressive disorder, bipolar affective disorder, and sporadic Creutzfeldt–Jakob disease (Lemos et al. 2009; Souza et al. 2010). This pipeline integrates databases freely available through the internet such as the Goldenpath-UCSC and NCBI. The alignment process was implemented and improved through the use of CLCbio Workbench Combined® software, together with the CUBE® hardware version 1.06. We performed multiple batches of Smith–Waterman alignments using an ESTs database downloaded from the Goldenpath and mRNA sequences downloaded from the NCBI. Such alignments represent the core of this whole process, and the following detection of SNPs and INDELs is based on it. The detection consists of a simple parsing of the alignments in order to determine the presence of variations, providing also positional details. In a latter process, the candidate sequences predicted are aligned with the NCBI SNPs database (<http://www.ncbi.nlm.nih.gov/snp/>) to search for previous deposited and confirm the identification of potential new SNPs and INDELs.

Several softwares used to predict pathogenicity of genetic variations have been developed. These computational methods are always debatable because of the often discrepancy between some of them, even when analyzing the same variation. Recently, Thusberg et al. (Thusberg et al. 2011) tested a set of over 40,000 variations known as pathogenic and neutral variants in nine softwares freely available at the internet, which are used to predict pathogenicity. The softwares evaluated were MutPred, nsSNPAnalyzer, Panther, PhD-SNP, PolyPhen, PolyPhen2, SIFT, SNAP, and SNPs&GO. The programs' performance ranged from poor, in the method nsSNPAnalyzer, to reasonably good in the methods SNPs&GO and MutPred.

Our group recently applied an analysis of prediction to better comprehend a missense substitution candidate for Fahr's disease at the MGEA6 gene and in a proline-rich and highly conserved protein domain. SAP, PolyPhen, and I Mutant suggest that this variation is predicted to cause



**Fig. 1** A pipeline suggested to merge different data resources in order to combine expression studies and genetic variations and predicted to detect potential new polymorphisms, including INDELs



probable molecular deleterious effects, but the PMUT considered this same mutation as neutral. The SIFT program generated conflicting findings, depending if we used the UniProt-SWISS PROT 56.6, NCBI, or UniProt-TrEMBL 39.6 databases (Lemos et al. 2011).

Several other approaches designed to detect SNPs and INDELs also have been developed, including TIARA and CHIP Bioinformatics Tools (Dereeper et al. 2011; Hong et al. 2011). Annotate Variation is a software freely available at <http://www.openbioinformatics.org/annovar> that deserves to be highlighted and allows to annotate SNP and INDEL, examining their functional consequence on genes, inferring cytogenetic bands, reporting functional importance scores, finding variants in conserved regions, and identifying variants reported in the 1000 Genomes Project and dbSNP (Wang et al. 2010). Thus, bioinformatics became an excellent new tool to integrate detection and functional inference methods derived from genomic studies, and here, we propose a new pipeline to merge different data resources in order combine expression and candidate variations predicted to detect potential new polymorphisms, including INDELs (Fig. 1).

## Perspectives

Relationships between INDELs and functional aspects are still unclear. However, some studies have been investigating these relationships of insertions/deletions variation and modification of proteins through observing the protein–protein interaction (Kang et al. 2011). The INDEL PDB software revealed 117,266 INDEL sites in 11,294 proteins. An interesting relationship between the length of these INDELs (1–25 bp) and loop has been indicating that there were many short INDELs/loops but very few longer INDELs/loops (Hsing and Cherkasoy 2008).

Recent approaches to localize INDELs highlighted a consensus that they might work as markers spread throughout the human genome, deriving from a single mutation event. Small INDELs can be identified across by multiplexing platforms and a simple dye-labeling electrophoretic approach as well as analysis with high-throughput technologies (Pereira et al. 2009a; Pereira et al. 2009b). Other applications are forensic studies that revealed 48 diallelic INDEL markers on genotyping forensic samples collected from different biological samples related to criminal cases. These INDELs also presented superior efficiency on samples with low copy number (Francez et al. 2011).

Our group is currently dealing with different approaches that might optimize sequencing and bioinformatic analyses of short INDELs to broaden our research capabilities of identifying those intriguing genetic variations. Reporting these pitfalls might help other groups that are planning to

search for such interesting genetic variation and call their attention for the high rate of false positives in experiments that lack careful checking of double strand sequencing. Hopefully, INDELs might become a new trend in association studies in neuropsychiatric genetics since so far the level of significant and positive associations with the standard SNPs reported presents limited predictive application.

**Acknowledgments** We are greatly indebted to José Luiz de Lima Filho and Henrique Castelletti. This study received financial support from the following Brazilian funding agencies and academic bureaus: LIKA-JIKA, PROPESQ-UFPE, CAPES, CNPq, and FACEPE.

## Reference

- Ahn SM, Kim TH, Lee S et al (2009) The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res* 19:1622–1629
- Armbruster D, Moser DA, Strobel A (2009) Serotonin transporter gene variation and stressful life events impact processing of fear and anxiety. *Int J Neuropsychopharmacol* 12(3):393–401
- Bentley DR, Balasubramanian S, Swerdlow HP et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59
- Bhangale TR, Rieder MJ, Livingston RJ, Nickerson DA (2005) Comprehensive identification and characterization of diallelic insertion–deletion polymorphisms in 330 human candidate genes. *Hum. Mol. Genet.* 1459–69
- Bhangale TR, Stephens M, Nickerson DA (2006) Automating resequencing-based detection of insertion–deletion polymorphisms. *Nat Genet* 38:1457–1462
- Bond J, Scott S, Hampshire DJ et al (2003) Protein-truncating mutations in ASPM cause variable reduction in brain size. *Am J Hum Genet* 73:1170–1177
- Capellari S, Strammiello R, Saverioni D, Kretschmar H, Parchi P (2011) Genetic Creutzfeldt–Jakob disease and fatal familial insomnia: insights into phenotypic variability and disease pathogenesis. *Acta Neuropathol* 121:21–37
- DePristo MA, Banks E, Poplin R et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498
- Dereeper A, Nicolas S, Le Cunff L (2011) SNIPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. *BMC Bioinforma* 5:12–134
- Francez PA, Rodrigues EMR, Velasco AM, Santos SEB (2011) Insertion–deletion polymorphisms—utilization on forensic analysis. *Int J Legal Med.* doi:10.1007/s00414-011-0588-z
- Gul A, Hassan MJ, Hussain S, Raza SI, Chishti MS, Ahmad W (2006) A novel deletion mutation in CENPJ gene in a Pakistani family with autosomal recessive primary microcephaly. *J Hum Genet* 51:760–764
- Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr Sci* 80:524–535
- Hong D, Park SS, Ju YS et al (2011) TIARA: a database for accurate analysis of multiple personal genomes based on cross-technology. *Nucleic Acids Res* 39(Database issue):D883–D888
- Hsing M, Cherkasoy A (2008) Indel PDB: a database of structural insertions and deletions derived from sequence alignments of closely related proteins. *BMC Bioinforma* 25:9–293

- Jing R, Bolshakov V, Flavell AJ (2007) The tagged microarray marker (TAM) method for high-throughput detection of single nucleotide and indel polymorphisms. *Nat Protoc* 2:168–177
- Kang L, Zhang N, Zhang Y et al (2011) Molecular characterization and identification of a novel polymorphism of 200 bp indel associated with age at first egg of the promoter region in chicken follicle-stimulating hormone receptor (FSHR) gene. *Mol Biol Rep* 38 (5):2903–2909
- Kidd JM, Cooper GM, Donahue WF et al (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56–64
- Kim JI, Ju YS, Park H, Kim S et al (2009) A highly annotated whole-genome sequence of a Korean individual. *Nature* 460:1011–1015
- Ku CS, Loy EY, Salim A, Pawitan Y, Chia KS (2010) The discovery of human genetic variations and their use as disease markers: past, present and future. *J Hum Genet* 55:403–415
- Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Leal GF, Roberts E, Silva EO, Costa SMR, Hampshire DJ, Woods CG (2003) A novel locus for autosomal recessive primary microcephaly maps to 13q12.2. *J Med Genet* 40:540–542
- Lemos RR, Castelletti CH, Lima Filho JL, Marques ET, Oliveira JRM (2009) In silico identification of new genetic variations as potential risk factors for Alzheimer's disease in a microarray-oriented simulation. *J Mol Neurosci* 39:242–247
- Lemos RR, Oliveira DF, Zatz M, Oliveira J (2011) Population and computational analysis of the mgea6 p521a variation as a risk factor for familial idiopathic basal ganglia calcification (Fahr's disease). *J Mol Neurosci* 43:333–336
- Levy S, Sutton G, Ng PC et al (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5(10):e254
- Ley TJ, Mardis ER, Ding L et al (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456:66–72
- Li H, Ruan J, Durbin R (2008a) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18:1851–1858
- Li R, Li Y, Kristiansen K, Wang J (2008b) SOAP: short oligonucleotide alignment program. *Bioinformatics* 24:713–714
- Mao G, Pan X, Zhu BB et al (2007) Identification and characterization of OGG1 mutation in patients with Alzheimer's disease. *Nucleic Acids Res* 35:2759–2766
- McKernan KJ, Peckham HE, Costa GL et al (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* 19:1525–1541
- Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* 16:1182–1190
- Mills RE, Pittard WS, Mullaney JM et al (2011) Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res* 21:830–839
- Mullaney JM, Mills RE, Pittard WS, Devine SE (2010) Small insertions and deletions (INDELs). *Hum Mol Genet* 19:131–136
- Mullikin JC, Hunt SE, Cole CG et al (2000) An SNP map of human chromosome 22. *Nature* 407:516–520
- Ng SB, Buckingham KJ, Lee C et al (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 42:30–35
- Nicholas AK, Swanson EA, Cox JJ et al (2009) The molecular landscape of ASPM mutations in primary microcephaly. *J Med Genet* 46:249–253
- Oliveira JRM, Otto PA, Vallada H et al (1998) Analysis of a novel functional polymorphism within the promoter region of the serotonin transporter gene (5-HTT) in Brazilian patients affected by bipolar disorder and schizophrenia. *Am J Med Genet* 81:225–227
- Oliveira JRM, Carvalho DR, Pontual D et al (2000) Analysis of the serotonin transporter polymorphism (5-HTTLPR) in Brazilian patients affected by dysthymia, major depression and bipolar disorder. *Mol Psychiatry* 5(4):348–349
- Pereira R, Phillip C, Alves C, Amorim A, Carracedo A, Gusmão L (2009a) Insertion/deletion polymorphisms: a multiplex assay and forensic applications. *Forensic Sci Int Genet Suppl* 2:513–515
- Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmão L (2009b) A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis* 30:3682–3690
- Pushkarev D, Neff NF, Quake S (2009) Single-molecule sequencing of an individual human genome. *Nat Biotechnol* 27:847–852
- Salathia N, Lee HN, Sangster TA et al (2007) Indel arrays: an affordable alternative for genotyping. *Plant J* 51(4):727–737
- Sawyer SL, Howell WM, Brookes AJ (2003) Scoring insertion-deletion polymorphisms by dynamic allele-specific hybridization. *Biotechniques* 35:292–296
- Schuster SC, Miller W, Ratan A et al (2010) Complete Khoisan and Bantu genomes from southern Africa. *Nature* 463:943–947
- Snyder N, Du J, Gerstein M (2010) Personal genome sequencing: current approaches and challenges. *Genes & Development* 24:423–443
- Souza MBR, Lemos RR, Da Cunha JE, Filho JL, Oliveira J (2010) Searching for new genetic risk factors for neuropsychiatric disorders in expression databases. *J Mol Neurosci* 41:193–197
- Syvänen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2:930–942
- Thusberg J, Olatubosun A, Vihinen M (2011) Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* 32(4):358–368
- Togsværød M, Werge TM, Tankó LB (2008) Association of a dopamine beta-hydroxylase gene variant with depression in elderly women possibly reflecting noradrenergic dysfunction. *J Affect Disord* 106(1–2):169–172
- Walsh T, McClellan JM, McCarthy SE et al (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–543
- Wang J, Wang W, Li R et al (2008) The diploid genome sequence of an Asian individual. *Nature* 456:60–65
- Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38(16):e164
- Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G (2002) Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet* 71:854–862
- Wheeler DA, Srinivasan M, Egholm M et al (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452:872–876
- Yu CE, Dawson G, Munson J et al (2002) Presence of large deletions in kindreds with autism. *Am J Hum Genet* 71:100–115

## Capítulo 4 – Artigo 3

Genome-wide association analysis combined with CSF and Plasma Biomarkers (AB-40 & AB-42) highlight new risk factors for Alzheimer's Disease



Manuscript Number: NSL-13-192

Title: Genome-wide association analysis combined with CSF and Plasma Biomarkers (AB-40 & AB-42) highlight new risk factors for Alzheimer's Disease

Article Type: Research Paper

Keywords: Alzheimer's Disease, Mild Cognitive Impairment, quantitative traits, Alzheimer's Disease Neuroimaging Initiative (ADNI), Biomarkers.

Corresponding Author: Dr Joao Ricardo Oliveira,

Corresponding Author's Institution: UFPE

First Author: Manuela B Souza

Order of Authors: Manuela B Souza; Gilderlanio S Araújo; Ivan G Costa Filho; Joao Ricardo Oliveira

**Abstract:** The deposition of the senile plaque is a common characteristic pathology in the cerebral aging and additional with neurofibrillary degeneration form the definitive findings for the diagnostic of Alzheimer's disease (AD), thus being considered potential biomarkers diagnostic. Related potential biomarkers and SNPs associated can be used to identify risk individuals for dementias. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. There are three subsets of participants: 380 individuals for A $\beta$ -42 in cerebrospinal fluid biomarkers, 654 individuals for A $\beta$ -42 and A $\beta$ -40 in plasma biomarkers for baseline and 538 individuals for plasma biomarkers collected one year late (M12). We analyzed the main effects of quantitative traits of single nucleotide polymorphisms (SNPs) in A $\beta$ -40 and A $\beta$ -42 biomarkers. Both association test and a linear regression method were used to investigate the main effects of SNPs f A $\beta$ -40 and A $\beta$ -42 under MCI and AD individuals. The p-values results were adjusted for multiple comparisons by the Bonferroni method and we focused on corrected p-values ( $p < 0.001$ ). The subsets show different SNPs association in diagnostic groups. We identified five SNPs as potential candidates, in the regions of the TOMM40, PAMR1, TRIM9 and CCDC112 genes and three SNPs in the non-coding region reached association significance. In addition to known candidate, gene TOMM40, the genes TRIM9 and PAMR1 are associated with dementias. We identified one novel gene candidate CCDC112 and a novel SNP (rs5904473) on the chromosome X were associated with Plasma biomarkers related to AD.

Suggested Reviewers: Mohammad Ikram  
m.a.ikram@erasmusmc.nl

Nick Fox

n.fox@ucl.ac.uk

John Hardy  
j.hardy@ucl.ac.uk

Kari Stefansson  
kstefans@decode.is

Forbs Porter  
fdporter@mail.nih.gov

Adam Fleisher  
adam.fleisher@bannerhealth.com

Ching-Chuan Liu  
liucc@mail.ncku.edu.tw

Christopher M. Norris  
cnorr2@uky.edu



## Federal University of Pernambuco Department of Neuropsychiatry

Recife, Jan 25<sup>th</sup>, 2012

**S.G. Waxman MD PhD**

Editor-in-Chief of the **Neuroscience Letters**

We are extremely happy to submit the manuscript entitled “**Genome-wide association analysis combined with CSF and Plasma Biomarkers (AB-40 & AB-42) highlight new risk factors for Alzheimer’s Disease**” to be considered for publication at this prestigious journal in a Original Articles format.

**This article reports SNPs associated with Alzheimer’s Disease in genes (PAMR1, TRIM9 and CCDC112) still not yet associated with this neurodegenerative condition. We believe this is an original finding with potential new contribution for the understanding of the genetic culprits in AD.**

**Financial Disclosures:** There is no conflict of interest in this study.

We believe that this analysis will be very well accepted by Neuroscience Letters.

We hope to hear from you soon with good news!

Best regards,

# Corresponding author: Dr. João Ricardo Mendes de Oliveira MD, PhD.

Department of Neuropsychiatry

Federal University of Pernambuco (UFPE), 50670901- Recife, Pernambuco – Brazil

e-mail: [joao.ricardo@ufpe.br](mailto:joao.ricardo@ufpe.br)

Author: Manuela Barbosa Rodrigues de Souza

Federal University of Pernambuco (UFPE), 50670901 Recife, PE, Brazil

Author: Gilderlanio Santana de Araújo

Federal University of Pernambuco (UFPE), 50670901 Recife, PE, Brazil

Author: Ivan Gesteira Costa Filho

Interdisciplinary Center for Clinical Research (IZKF), University Hospital RWTH, Aachen, Germany.

## Highlights

- We used Alzheimer's Disease Neuroimaging Initiative (ADNI) database.
- We analyzed the main effects of quantitative traits of single nucleotide polymorphisms (SNPs) in  $A\beta$ -40 and  $A\beta$ -42 biomarkers.
- We identified five SNPs as potential candidates and three SNPs in the non-coding region reached association significance.
- We identified one novel gene associated with Plasma biomarkers related to AD.

**Genome-wide association analysis combined with CSF and Plasma Biomarkers (AB-40 & AB-42) highlight new risk factors for Alzheimer's Disease**

Souza MBR<sup>1+</sup>, Araújo GS<sup>2+</sup>, Costa IG<sup>2,3</sup>, Oliveira JRM<sup>1,4#</sup>, for the Alzheimer's Disease Neuroimaging Initiative\*

1- Keizo Asami Laboratory (LIKA) – Federal University of Pernambuco, Recife-PE, Brazil.

Author: Manuela Barbosa Rodrigues de Souza

Federal University of Pernambuco (UFPE), 50670901 Recife, PE, Brazil

2- Center of Informatics - Federal University of Pernambuco, Recife-PE, Brazil.

Author: Gilderlanio Santana de Araújo

Federal University of Pernambuco (UFPE), 50670901 Recife, PE, Brazil

3 – Interdisciplinary Center for Clinical Research (IZKF), University Hospital RWTH, Aachen, Germany.

Author: Ivan Gesteira Costa Filho

University Hospital RWTH Aachen, 52074, Aachen, Germany,

4 - Neuropsychiatry Department - Federal University of Pernambuco, Recife-PE, Brazil.

# Corresponding author: João Ricardo Mendes de Oliveira

e-mail: joao.ricardo@ufpe.br

Department of Neuropsychiatry

Av. Prof. Moraes Rego, s/n - cidade Universitária.

Telephone: (81) 21268539 Fax: (81) 21268539

Federal University of Pernambuco (UFPE), 50670901 Recife, PE, Brazil

<sup>+</sup> These authors contributed equally to this work.

\* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.ucla.edu](http://adni.loni.ucla.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.ucla.edu/wpcontent/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

## ABSTRACT

The deposition of the senile plaque is a common characteristic pathology in the cerebral aging and additional with neurofibrillary degeneration form the definitive findings for the diagnostic of Alzheimer's disease (AD), thus being considered potential biomarkers diagnostic. Related potential biomarkers and SNPs associated can be used to identify risk individuals for dementias. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. There are three subsets of participants: 380 individuals for  $A\beta$ -42 in cerebrospinal fluid biomarkers, 654 individuals for  $A\beta$ -42 and  $A\beta$ -40 in plasma biomarkers for baseline and 538 individuals for plasma biomarkers collected one year late (M12). We analyzed the main effects of quantitative traits of single nucleotide polymorphisms (SNPs) in  $A\beta$ -40 and  $A\beta$ -42 biomarkers. Both association test and a linear regression method were used to investigate the main effects of SNPs f  $A\beta$ -40 and  $A\beta$ -42 under MCI and AD individuals. The p-values results were adjusted for multiple comparisons by the Bonferroni method and we focused on corrected p-values ( $p < 0.001$ ). The subsets show different SNPs association in diagnostic groups. We identified five SNPs as potential candidates, in the regions of the TOMM40, PAMR1, TRIM9 and CCDC112 genes and three SNPs in the non-coding region reached association significance. In addition to known candidate, gene TOMM40, the genes TRIM9 and PAMR1 are associated with dementias. We identified one novel gene candidate CCDC112 and a novel SNP (rs5904473) on the chromosome X were associated with Plasma biomarkers related to AD.

**Keywords:** Alzheimer's Disease, Mild Cognitive Impairment, quantitative traits, Alzheimer's Disease Neuroimaging Initiative (ADNI), Biomarkers.



## INTRODUCTION

Genome-wide association studies have helped to further our understanding of the pathogenic mechanism of several complex and multifactorial disorders such neurodegenerative disease. Neurodegenerative diseases are progressive diseases characterized by accumulation of specific protein and loss of neuronal followed by cognitive impairments and memory deficiencies [20].

Alzheimer's disease (AD) is the most common progressive neurodegenerative disease and represents a major cause of disability for elderly patients with dementia in worldwide [23]. AD can be divided into three phases, an initial pre-symptomatic phase, a prodromal stage known as Mild Cognitive Impairment (MCI) and a third phase when patients show dementia with impairments in multiple domains and loss of function in daily activities [31].

Quantitative traits are an alternative to foster the progress in elucidating genetics variants with strong phenotypic effects in complex diseases such as Late Onset Alzheimer's Disease (LOAD) and MCI [5]. Many of these traits are highly heritable with genetic factors presenting high heritability in LOAD (58%-79%), however, the elucidation of their genetic architecture has been limited [8].

The abnormal deposition of amyloid- $\beta$  ( $A\beta$ ) protein in the brain plays an important role in AD, produces senile plaques containing, dystrophic neurites, neurofibrillary tangles (NFT), and neuropil threads, which constitute the common pathological hallmarks of AD, being considered potential clinical biomarker [3, 30]. The  $A\beta$  is detectable in cerebrospinal fluid (CSF) and plasma and is responsible to form a major component of senile plaques deposited in brain and have been used as biomarkers for diagnosis [9]. Changes in CSF and plasma levels of the  $A\beta$ -40 and  $A\beta$ -42 are potential diagnostic biomarkers for AD [9, 3]. Baseline plasma levels of  $A\beta$ -42



are higher in patients with AD and predict high risk of progression to dementia in normal individuals [30]. The discovery of additional genes that may be related to amyloid pathophysiology in AD and MCI allows the introduction of novel and potentially useful genetic markers for the presence and progression of AD pathology [17, 18].

In this paper we correlate the  $A\beta$ -40 and  $A\beta$ -42 levels in patients with AD and MCI genotyped with Illumina Human 610 – Quad Bead Chip platform, originated from the Alzheimer’s ADNI. This approach aims to try to correlate genetic polymorphisms with biomarker of disease activity risk and clinical profile.

## METHODS

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database ([adni.loni.ucla.edu](http://adni.loni.ucla.edu)). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5- year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private

corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years.” For up-to-date information, see [www.adni-info.org](http://www.adni-info.org) [13].

In this study, three subsets of biomarkers were used to examine the effects of each SNP on the ADNI cohort. These subsets are summarized in Table 1. In the first subset, we included 380 individuals for  $A\beta$ -42 in CSF; in the second subset, we included 654 individuals for  $A\beta$ -42 and  $A\beta$ -40 in plasma for baseline diagnostic and in the last subset were included 538 observed individuals for plasma biomarkers collected one year late (M12) from the ADNI cohort. The data met all detailed quality control of data previously reported for CSF [26] and for plasma biomarkers [27]. To reduce the likelihood of population stratification effects in the GWAS all individuals in this study are non-Hispanic Caucasian.

After restriction of the ADNI participants remained in the present study 380 individuals for CSF biomarkers investigation; 654 individuals for investigation of baseline plasma biomarkers levels and for M12 (one year late) plasma biomarkers levels remained 638 individuals distributed over the diagnostic groups (NC, MCI and AD). The distribution of individuals on diagnostic groups is presented on Table 1 and a statistical summary for each experimental design. Due to population stratification, the obtained genomic inflation factors to investigate CSF and Plasma biomarkers associations indicated low risk of confounding and the values for the subset is presented on Table 1 [27].

The Illumina Human 610-Quad BeadChip allows geneticists to investigate more than 610,000 SNPs, however this study was restricted to SNPs with minor allele frequency (MAF) greater than 0.2 and SNPs with 'rsIDs'. The MAF value was used to exclude rare SNPs, in consequence of the smaller size of samples and aims to reduce both the false-positive results and severity of the multiple correction tests. The MAF value is stringent, so the number of final SNPs for analysis varies for each experimental design and can be viewed on Table 1.

We performed both a standard allelic test and a linear regression model for quantitative trait association, using the PLINK software [12], release v1.07 [22]. Both tests were applied to each GWAS analysis to calculate the main effects of the SNPs on the target quantitative biomarkers variables. The association test and linear regression results were obtained with the confidence interval of 0.95 and p-values were adjusted to Bonferroni correction. The statistical results were plotted using Manhattan plot and Quantile-Quantile (Q-Q) plot and are presented on Figure 1.

Next, we used the GeneMANIA [14] to perform a network interaction analysis with genes associated to the biomarkers as query [32]. The interaction data arises from large-scale biological data sources as protein-protein interaction, protein domain similarity, pathways and gene expression profiles [32]. GeneMANIA indicates other genes, which are in the vicinity of the query network, therefore giving hints of possible biological mechanisms of the associated genes on the biomarkers or the disease. We base our analysis on the network setting the vicinity of genes to 20.

## RESULTS

In the subsequent GWAS analyses, after the Bonferroni correction (p-value < 0.001), using an additive linear regression model test, we showed that all biomarkers subsets differ between AD and MCI groups. Seven SNPs achieved significance after the

correction. The Figure 1 displays Manhattan and Q-Q plot of  $A\beta$ -42 CSF baseline diagnosis;  $A\beta$ -42 plasma baseline diagnosis;  $A\beta$ -42 plasma M12 and  $A\beta$ -40 plasma M12.

We focused in the SNPs with adjusted p-values and summarize the top SNPs in the Table 2. We found two SNPs significantly associated with CSF biomarker levels baseline. In baseline plasma AD diagnostic group only one SNP rs13395562 on non-coding region was significantly associated. In M12 AD diagnostic group five SNPs significantly associated with CSF biomarker level of  $A\beta$ -40 and  $A\beta$ -42.

In addition to GWAS experiment, we investigate the biomarkers levels in function of genotypes. The Figure 2 presents statistics (mean and standard error) for  $A\beta$ -40 and  $A\beta$ -42 biomarkers level as a function of genotype over all diagnostic groups for the most significant associated SNPs.  $A\beta$ -42 CSF Baseline levels in diagnostic groups (AD, MCI and NC) stratified by rs2075650 genotype showed significant differences between GG homozygous and GA heterozygous. However,  $A\beta$ -42 CSF Baseline stratified by rs49200 genotype, such as  $A\beta$ -42 and  $A\beta$ -42 Plasma Baseline and M12, showed linear trend in mean levels within and across all diagnostic groups for the genotypes analyzed. Table 3 shows the GA genotype of rs2075650, the GG genotype of rs13395562 and AA genotyped of rs644755, rs10501140 and rs11950355 presented at lower frequencies in all diagnostic groups, while the GG genotype of rs17802296 and genotype AA rs5904473 presented at high frequencies in all diagnostic groups. Already rs4900200 genotypic frequency showed similar frequencies between the three subgroups. The possible role of these SNPs in AD and MCI pathogenesis need detailed investigation.

First, we performed GeneMANIA predictions based on a query with genes *TOMM40*, *PAMR1*, *TRIM9* and *CCDC112*. This indicated  $A\beta$  as one of the top 20



interacting genes (see from the supplementary material) and all query genes were connected with a small network distance, which indicate tight interactions between them. We also repeated the same analysis including APP in the query network. Now, APOE gene is indicated as an important interaction partner. Moreover, two proteins previously related to Alzheimer, GAPDH and SNPA25 support the interaction of TRIM9 and CCDC112 to A $\beta$  [6, 1].

## DISCUSSION

In this paper, we investigated plasma and CSF biomarkers CSF A $\beta$ -42, Plasma A $\beta$ -40 and A $\beta$ -42 as quantitative traits to discover genetic variant associations using genotype data obtained from the ADNI cohort. The use of phenotypes as quantitative traits in GWAS have been successfully used to test association between biomarkers and SNPs. Additionally the biomarkers have been shown an increased power over case-control designs [10, 17, 21].

The *TOMM40*, a translocase of outer mitochondrial membrane 40 homolog, is essential for protein import into mitochondria, the aging decreases the number of mitochondria and also increases risk of developing AD [15]. *TOMM40* alleles have been associated to an increased risk of developing late-onset AD [7, 25]. *TOMM40* is an adjacent and approximately 15 kb upstream to *APOE*, established as one of the susceptibility gene for LOAD [4]. *TOMM40* variable-length poly-T sequence polymorphism (rs10524523) in combination with *APOE* alleles (E2, E3, E4) has been reported as influences in the LOAD [24]. Recently, was analyzed the impact of *TOMM40* poly-T in combination with *APOE* alleles (E2, E3, E4) on LOAD incidence in a group of 414 LOAD patients, 173 centenarians and 305 neurologically healthy individuals. They, demonstrated for some analyses poly-T affecting LOAD risk [19].

Using quantitative trait in a GWAS, the *TOMM40* gene, specifically the SNP rs2075650, has been identified as putative susceptible loci and genetic variants associated with AD [10]. However, used CSF biomarkers ( $A\beta$ -42, t-tau, p-tau181p, p-tau181p/ $A\beta$ -42, and t-tau/ $A\beta$ -42) data ADNI shows rs2075650 (*TOMM40*) associated with  $A\beta$ -42, p-tau181p/ $A\beta$ -42 and t-tau/ $A\beta$ -42 [17]. An intron SNP rs2075650 of gene *TOMM40* is particularly associated with  $A\beta$ -42 in AD and MCI subjects. This SNP was previously described with a positive association with *APOE* biomarker as quantitative traits [21].

Few studies have correlated the *PAMR1*, peptidase domain containing associated with muscle regeneration 1, gene with AD, suggesting now a potentially new gene involved with this neurodegenerative disease. *PAMR1* has been studied in microarrays expression analysis of post mortem samples from patients with late onset AD and matched controls (LOAD, final 187 controls, 176 cases). Another analysis was done with induced stem cells from sporadic and familial cases of AD [33, 16].

According to the Allen Institute Human Brain atlas, the *PAMR1* presents a selective pattern of expression along the hippocampus, a known region involved in AD. The hippocampal formation is mainly down regulated for this gene, however, segments of the CA1 region and parahippocampal formation are up regulated [11].

TRIPartite Motif (TRIM) family proteins are ring finger domain-containing. Member of the family of TRIM proteins, *TRIM9* tripartite motif containing 9, location in 14q22.1 region is highly expressed in the cerebral cortex, in the Purkinje cells of the cerebellum and in the hippocampus; has neuronal functions associated with diseases neurodegenerative. Previous study showed *TRIM9* severely decreased in the affected brain areas in Parkinson's disease and dementia with Lewy bodies [2, 28, 29].

The association test also allowed us to identify one novel candidate gene, *CCDC112*, and the novel SNP rs5904473 in non-coding region of the chromosome X (Table 2). We found strong association for the SNP rs11950355, located in the intron region in gene *CCDC112*, with  $A\beta$ -42 plasma biomarker in AD subset. The gene *CCDC112*, coiled-coil domain containing 112, is located at 5q22.3, has two isoforms (isoform 1 and isoform 2), but the isoform 1 is higher expressed than isoform 2. *CCDC112* has not been previously reported association between with AD or neurodegenerative diseases.

## CONCLUSION

Our analysis of the ADNI GWAS identified several putative loci that are in genetic association with  $A\beta$ -42 levels in CSF and  $A\beta$ -40 levels in CSF and Plasma. The most significant associations were identified using AD subjects probably because the variances of the CSF biomarker levels are smaller in MCI and normal subjects. Thus, our analysis demonstrates that quantitative trait observing biomarkers levels of  $A\beta$ -40 and  $A\beta$ -42 by genome-wide screening can reveal additional insights into the mechanism that connects these biomarkers with potentially new candidate genes for AD and MCI.

The SNP rs11950355, located in the intron region of gene *CCDC112* and rs5904473 in a non-coding region were significantly associated with  $A\beta$ -42 in AD subjects. These warrant detailed investigation regarding possible role with AD.

Replication studies with independent, larger samples will be important to confirm these findings. As well as, assess if a panel of genetic markers can be combined with CSF and Plasma analyses to better predict longitudinal outcomes or response to emergent therapeutics. In future studies, it will also be important include looking at the interaction of SNP and diagnosis and gene or pathway-based analyses to further investigate associations with biomarker levels.



**Financial Disclosures:** There is no conflict of interest in this study.

## ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorphix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

This study received financial support from the following academic bureaus and Brazilian funding agencies: Centro de Informática - CIN, LIKA-JIKA, UFPE, FACEPE and CNPq.

## LEGENDS

Figure 1. Manhattan and Q-Q plots of the quantitative trait association analysis in AD and MCI diagnostic groups. **A**,  $A\beta$ -42 CSF baseline AD diagnosis; **B**,  $A\beta$ -42 CSF baseline MCI diagnosis; **C**,  $A\beta$ -42 Plasma Baseline AD diagnosis; **D**,  $A\beta$ -42 Plasma M12 AD diagnosis; **E**,  $A\beta$ -40 Plasma M12 AD diagnosis.

Figure 2. Mean of biomarker levels as a function of CSF and plasma baseline and plasma M12 genotypes over all diagnostic groups (NC, MCI and AD) for the most significant associated SNPs.

Table1. Details of experimental designs performed.

Abbreviations: CSF, cerebrospinal fluid; Chr, chromosome; M12, one year late; AD, Alzheimer Disease; MCI, Mild Cognitive Impairment; SNP, single nucleotide

polymorphism; MAF, minor allele frequency; SD, standard deviation; L, level; A, age; N, number.

Table 2. Top SNPs with  $p$ -value  $< 10^{-4}$ . SNPs associated with CSF and Plasma biomarkers in diagnostic groups (NC, MCI and AD).

Abbreviations: CSF, Cerebrospinal fluid; Chr, chromosome; M12, one year late; AD, Alzheimer Disease; MCI, Mild Cognitive Impairment; NC, normal control; SNP, single nucleotide polymorphism.

Table 3. Genotypic frequency.

Abbreviations: BL, baseline; M12, one year late; AD, Alzheimer Disease; MCI, Mild Cognitive Impairment; NC, normal control; N, number.

## REFERENCES

- [1] Bailey, J.A., Lahiri, D.K., 2006. Neuronal differentiation is accompanied by increased levels of SNAP-25 protein in fetal rat primary cortical neurons: implications in neuronal plasticity and Alzheimer's disease. *Ann N Y Acad Sci* 1086, 54-65.
- [2] Berti, C., Messali, S., Ballabio, A., Reymond, A., Meroni, G., 2002. TRIM9 is specifically expressed in the embryonic and adult nervous system. *Mech Dev* 113, 159-62.
- [3] Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 82, 239-259.
- [4] Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., Pericak-Vance, M.A., 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261, 921-923.
- [5] Cruchaga, C., Kauwe, J.S., Mayo, K., Spiegel, N., Bertelsen, S., Nowotny, P., Shah, A.R., Abraham, R., Hollingworth, P., Harold, D., Owen, M.M., Williams, J., Lovestone, S., Peskind, E.R., Li, G., Leverenz, J.B., Galasko, D., Alzheimer's Disease Neuroimaging Initiative., Morris, J.C., Fagan, A.M., Holtzman, D.M., Goate, A.M., 2010. SNPs associated with cerebrospinal fluid phospho-tau levels influence rate of decline in Alzheimer's disease. *PLoS Genet* 6, e1001101.
- [6] Cumming, R.C., Schubert, D., 2005. Amyloid-beta induces disulfide bonding and aggregation of GAPDH in Alzheimer's disease. *FASEB J* 19, 2060-2062.
- [7] Devi, L., Prabhu, B.M., Galati, D.F., Avadhani, N.G., Anandatheerthavarada, H.K., 2006. Accumulation of amyloid precursor protein in the mitochondrial import channels of human Alzheimer's disease brain is associated with mitochondrial dysfunction. *J Neurosci* 26, 9057-9068.

- [8] Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., Fiske, A., Pedersen, N.L., 2006. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 63, 168-174.
- [9] Graff-Radford, N.R., Crook, J.E., Lucas, J., Boeve, B.F., Knopman, D.S., Ivnik, R.J., Smith, G.E., Younkin, L.H., Petersen, R.C., Younkin, S.G., 2007. Association of low plasma A $\beta$ 42/A $\beta$ 40 ratios with increased imminent risk for mild cognitive impairment and Alzheimer disease. *Arch Neurol* 64, 354-362.
- [10] Han, M.R., Schellenberg, G.D., Wang, L.S., Alzheimer's Disease Neuroimaging Initiative., 2010. Genome-wide association reveals genetic effects on human A $\beta$ 42 and  $\tau$  protein levels in cerebrospinal fluids: a case control study. *BMC Neurol* 10, 90.
- [11] <http://human.brain-map.org/> (accessed 17/07/2012).
- [12] <http://pngu.mgh.harvard.edu/~purcell/plink/> (accessed 04/06/2012).
- [13] <http://www.adni-info.org> (accessed 23/03/2012).
- [14] <http://www.genemania.org> (accessed 16/07/2012).
- [15] Humphries, A.D., Streimann, I.C., Stojanovski, D., Johnston, A.J., Yano, M., Hoogenraad, N.J., Ryan, M.T., 2005. Dissection of the mitochondrial import and assembly pathway for human Tom40. *J Biol Chem* 280, 11535-11543.
- [16] Israel, M.A., Yuan, S.H., Bardy, C., Reyna, S.M., Mu, Y., Herrera, C., Hefferan, M.P., Van Gorp, S., Nazor, K.L., Boscolo, F.S., Carson, C.T., Laurent, L.C., Marsala, M., Gage, F.H., Remes, A.M., Koo, E.H., Goldstein, L.S., 2012. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* 482, 216-220.
- [17] Kim, S., Swaminathan, S., Shen, L., Risacher, S.L., Nho, K., Foroud, T., Shaw, L.M., Trojanowski, J.Q., Potkin, S.G., Huettel, M.J., Craig, D.W., DeChairo, B.M., Aisen, P.S., Petersen, R.C., Weiner, M.W., Saykin, A.J., Alzheimer's Disease Neuroimaging Initiative., 2011. Genome-wide association study of CSF biomarkers A $\beta$ 1-42, t-tau, and p-tau181p in the ADNI cohort. *Neurology* 76, 69-79.
- [18] Kölsch, H., Jessen, F., Wiltfang, J., Lewczuk, P., Dichgans, M., Kornhuber, J., Frölich, L., Heuser, I., Peters, O., Schulz, J.B., Schwab, S.G., Maier, W., 2008. Influence of SORL1 gene variants: Association with CSF amyloid-beta products in probable Alzheimer's disease. *Neurosci Lett* 440, 68-71.
- [19] Maruszak, A., Pepłońska, B., Safranow, K., Chodakowska-Żebrowska, M., Barcikowska, M., Zekanowski, C., 2012. TOMM40 rs10524523 polymorphism's role in late-onset Alzheimer's disease and in longevity. *J Alzheimers Dis* 28, 309-322.
- [20] Nath, S., Agholme, L., Kurudenkandy, F.R., Granseth, B., Marcusson, J., Hallbeck, M., 2012. Spreading of Neurodegenerative Pathology via Neuron-to-Neuron Transmission of A $\beta$  Amyloid. *J Neurosci* 32, 8767-8777.



- [21] Potkin, S.G., Guffanti, G., Lakatos, A., Turner, J.A., Kruggel, F., Fallon, J.H., Saykin, A.J., Orro, A., Lupoli, S., Salvi, E., Weiner, M., Macciardi, F., Alzheimer's Disease Neuroimaging Initiative., 2009. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. *PLoS One* 4, e6501.
- [22] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559–575.
- [23] Rosenberg, P.B., Lyketsos, C., 2008. Mild cognitive impairment: searching for the prodrome of Alzheimer's disease. *World Psychiatry* 7, 72-78.
- [24] Roses, A.D., 2011. Genetic and clinical implications of a variable poly-T repeat within the *TOMM40* gene for late-onset Alzheimer's disease. *Arch Neurol* 67, 536–541.
- [25] Roses, A.D., Lutz, M.W., Amrine-Madsen, H., Saunders, A.M., Crenshaw, D.G., Sundseth, S.S., Huentelman, M.J., Welsh-Bohmer, K.A., Reiman, E.M., 2010. A *TOMM40* variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J* 10, 375-384.
- [26] Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M., Trojanowski, J.Q., Alzheimer's Disease Neuroimaging Initiative., 2009. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* 65, 403-413.
- [27] Soares, H.D., Potter, W.Z., Pickering, E., Kuhn, M., Immermann, F.W., Shera, D.M., Ferm, M., Dean, R.A., Simon, A.J., Swenson, F., Siuciak, J.A., Kaplow, J., Thambisetty, M., Zagouras, P., Koroshetz, W.J., Wan, H.I., Trojanowski, J.Q., Shaw, L.M. for the Biomarkers Consortium Alzheimer's Disease Plasma Proteomics Project., 2012. Plasma Biomarkers Associated With the Apolipoprotein E Genotype and Alzheimer Disease. *Arch Neurol* 16, 1-8.
- [28] Song, S., Ge, Q., Wang, J., Chen, H., Tang, S., Bi, J., Li, X., Xie, Q., Huang, X., 2011. TRIM-9 functions in the UNC-6/UNC 40 pathway to regulate ventral guidance. *J Genet Genomics* 38, 1-11.
- [29] Tanji, K., Kamitani, T., Mori, F., Kakita, A., Takahashi, H., Wakabayashi, K., 2010. TRIM9, a novel brain-specific E3 ubiquitin ligase, is repressed in the brain of Parkinson's disease and dementia with Lewy bodies. *Neurobiol Dis* 38,210-218.
- [30] Tarawneh, R., Holtzman, D.M., 2010. Biomarkers in translational research of Alzheimer's disease. *Neuropharmacology* 59, 310-322.
- [31] Trojanowski, J.Q., Vandeerstichele, H., Korecka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W.Z., Weiner, M.W., Jack, C.R. Jr., Jagust, W., Toga, A.W., Lee, V.M., Shaw, L.M., Alzheimer's Disease Neuroimaging Initiative., 2010. Update on the biomarker

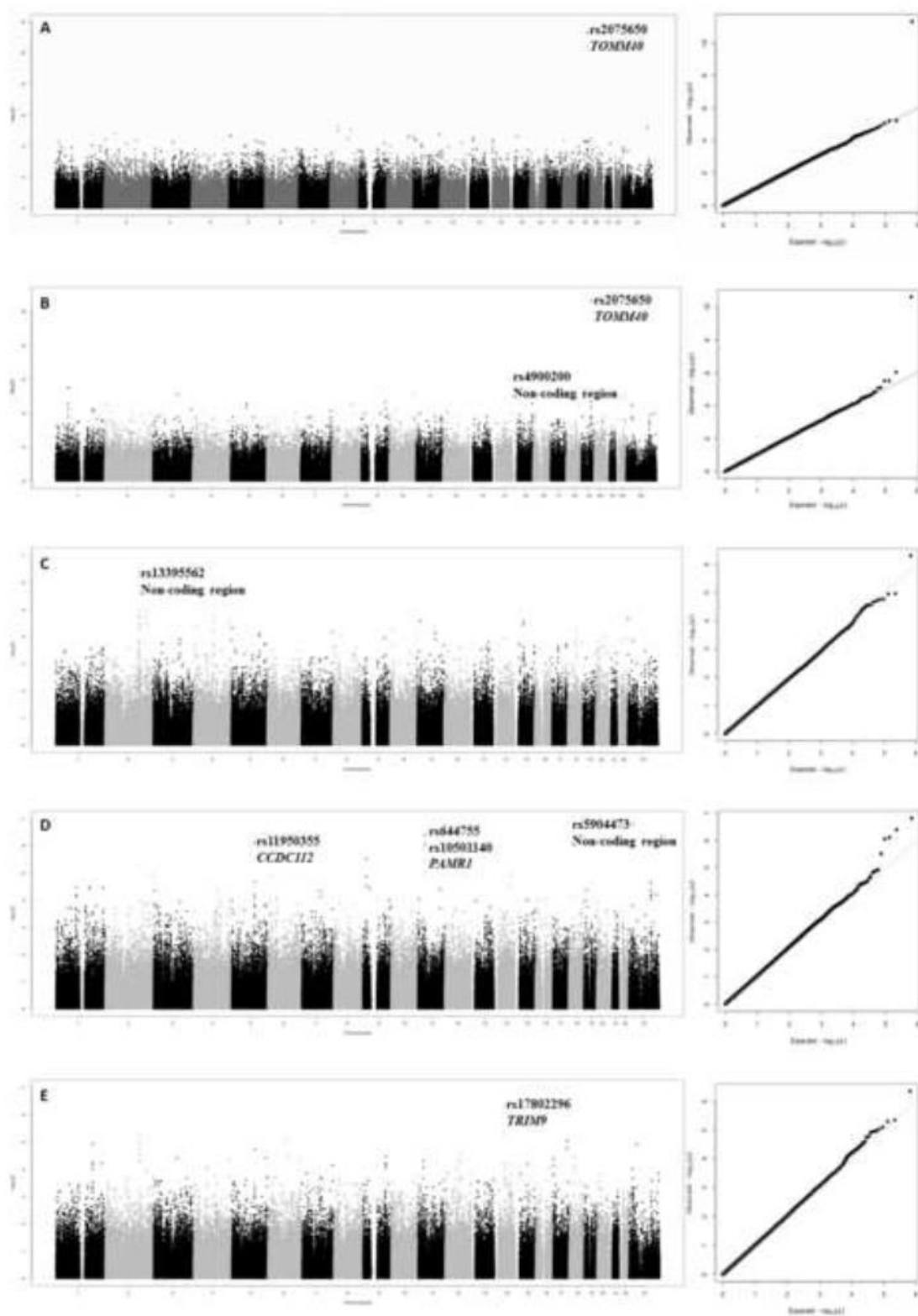
core of the Alzheimer's Disease Neuroimaging Initiative subjects. *Alzheimers Dement* 6, 230-238.

[32] Warde-Farley, D., Donalson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G.D., Morris, Q., 2010. The GeneMania prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res (Web Server issue)*, W214-220.

[33] Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., McCorquodale, D.S., Cuello, C., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Joshipura, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J.V., NACC-Neuropathology Group., Heward, C.B., Reiman, E.M., Stephan, D., Hardy, J., Myers, A.J., 2009. Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* 84, 445-458.

Figure1.tif

[Click here to download high resolution image](#)



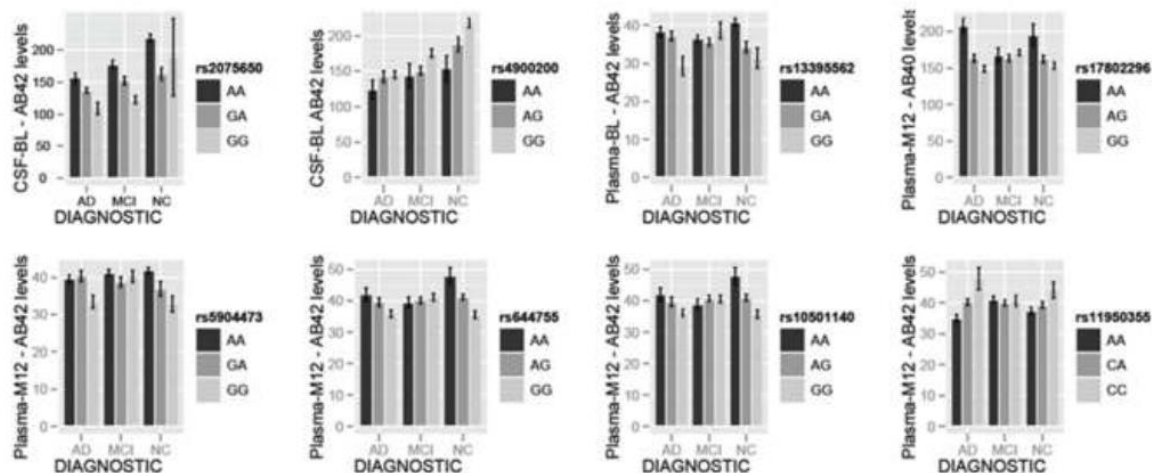


Table 1.docx

| CSF-Baseline               |               |                 |                                                                       |                                                                     |                          |                                                                                                    |                                                                   |
|----------------------------|---------------|-----------------|-----------------------------------------------------------------------|---------------------------------------------------------------------|--------------------------|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|
| Design Experimental        | Genotype rate | SNPs (MAF >0.2) | AB-40 (Mean   SD)                                                     | AB42 (Mean   SD)                                                    | Genomic inflation factor | Individuals                                                                                        | Age                                                               |
| NC-MCI                     | 0.983615      | 328.133         | L <sub>NC</sub> = 205.67   55.72<br>L <sub>MCI</sub> = 163.00   55.41 | --                                                                  | 1.03346                  | N <sub>NC</sub> = 107; Male = 56; Female = 51<br>N <sub>MCI</sub> = 179; Male = 120; Female = 59   | A <sub>NC</sub> = 75.44   5.16<br>A <sub>MCI</sub> = 74.8   7.62  |
| NC-AD                      | 0.983143      | 327.433         | L <sub>AD</sub> = 142.82   39.56                                      |                                                                     | 1.06704                  | N <sub>AD</sub> = 94; Male = 54; Female = 40                                                       | A <sub>AD</sub> = 75.08   8.02                                    |
| Plasma-Baseline            |               |                 |                                                                       |                                                                     |                          |                                                                                                    |                                                                   |
| NC-MCI                     | 0.983377      | 3283.62         | L <sub>NC</sub> = 147.57   51.86<br>L <sub>MCI</sub> = 150.62   56.43 | L <sub>NC</sub> = 37.19   11.66<br>L <sub>MCI</sub> = 36.26   12.54 | 1                        | N <sub>NC</sub> = 192; Male = 106; Female = 86<br>N <sub>MCI</sub> = 314; Male = 209; Female = 105 | A <sub>NC</sub> = 75.75   4.95<br>A <sub>MCI</sub> = 75.29   7.19 |
| NC-AD                      | 0.983907      | 328.946         | L <sub>AD</sub> = 151.25   47.58                                      | L <sub>AD</sub> = 36.43   10.50                                     | 1                        | N <sub>AD</sub> = 148; Male = 80; Female = 68                                                      | A <sub>AD</sub> = 75.54   7.49                                    |
| Plasma-M12 (one year late) |               |                 |                                                                       |                                                                     |                          |                                                                                                    |                                                                   |
| NC-MCI                     | 0.983769      | 328.608         | L <sub>NC</sub> = 159.26   49.43<br>L <sub>MCI</sub> = 168.02   45.28 | L <sub>NC</sub> = 39.04   11.12<br>L <sub>MCI</sub> = 40.60   11.00 | 1.05784                  | N <sub>NC</sub> = 191; Male = 107; Female = 84<br>N <sub>MCI</sub> = 250; Male = 168; Female = 82  | A <sub>NC</sub> = 76.69   5.31<br>A <sub>MCI</sub> = 76.79   7.17 |
| NC-AD                      | 0.983635      | 327.721         | L <sub>AD</sub> = 157.18   48.72                                      | L <sub>AD</sub> = 38.06   11.67                                     | 1.00948                  | N <sub>AD</sub> = 197; Male = 112; Female = 85                                                     | A <sub>AD</sub> = 76.37   7.36                                    |

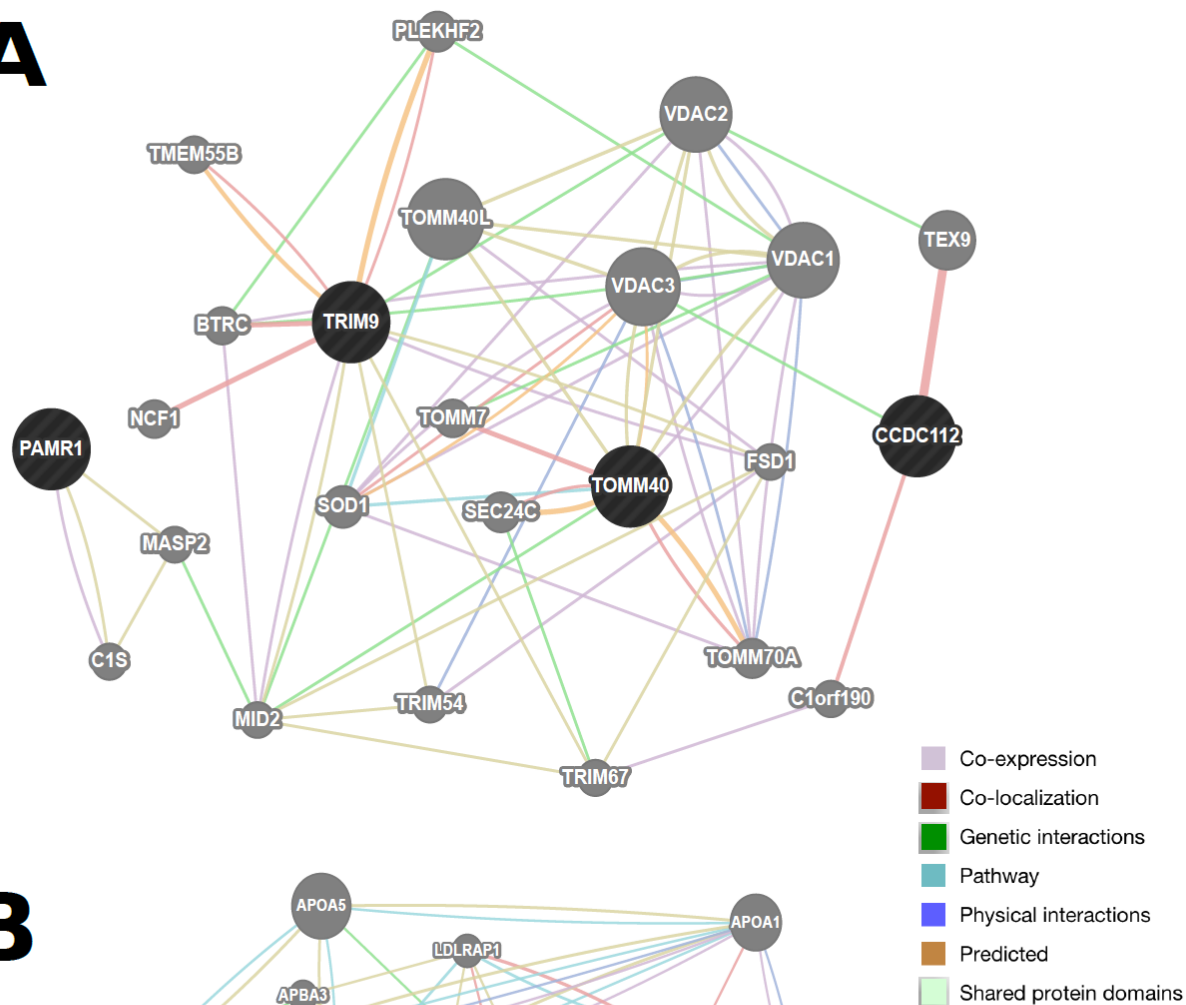


| Table 2.doc CSF Baseline |     |            |                   |                      |                |             |
|--------------------------|-----|------------|-------------------|----------------------|----------------|-------------|
| Diagnostic               | Chr | SNP        | Gene              | Associated Biomarker | p-value        | Test        |
| AD                       | 19  | rs2075650  | <i>TOMM40</i>     | AB-42                | 0,000000004712 | Association |
| MCI                      | 19  | rs2075650  | <i>TOMM40</i>     | AB-42                | 0,00000002332  | Association |
| MCI                      | 14  | rs4900200  | Non-coding region | AB-42                | 0,0009224      | Association |
| Plasma Baseline          |     |            |                   |                      |                |             |
| AD                       | 2   | rs13395562 | Non-coding region | AB-42                | 0,0000004971   | Linear      |
| Plasma M12               |     |            |                   |                      |                |             |
| AD                       | 11  | rs644755   | <i>PAMR1</i>      | AB-42                | 0,0003998      | Linear      |
| AD                       | 11  | rs10501140 | <i>PAMR1</i>      | AB-42                | 0,0008856      | Linear      |
| AD                       | 5   | rs11950355 | <i>CCDC112</i>    | AB-42                | 0,0007763      | Linear      |
| AD                       | X   | rs5904473  | Non-coding region | AB-42                | 0.0000001528   | Linear      |
| AD                       | 14  | rs17802296 | <i>TRIM9</i>      | AB-40                | 0,0004605      | Linear      |

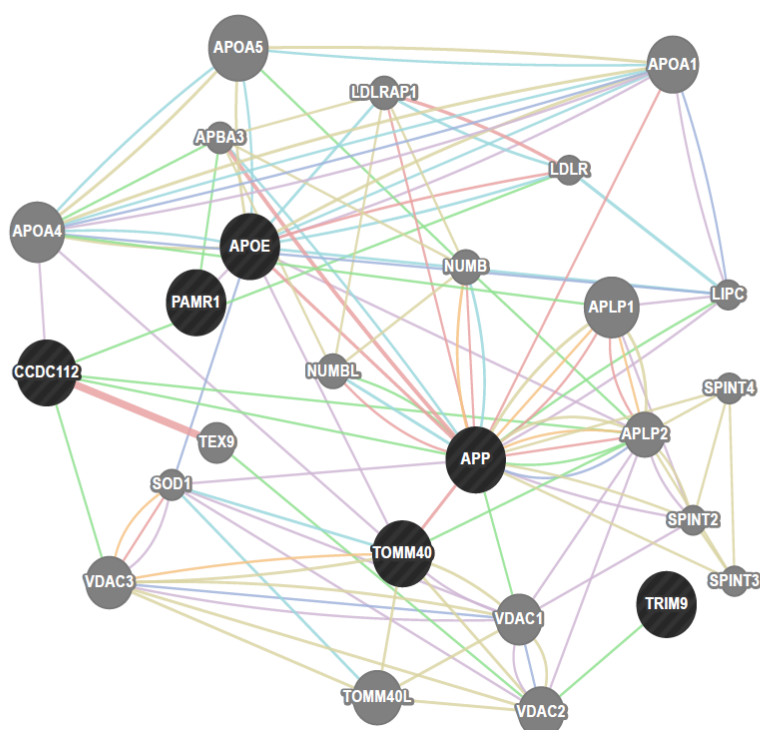
Table 3.doc

|            | AB42-CSF-BL |    |           |    | AB42-PLASMA-BL |     | AB40-PLASMA-M12 |     | AB42-PLASMA-M12 |     |          |     |            |     |            |     |
|------------|-------------|----|-----------|----|----------------|-----|-----------------|-----|-----------------|-----|----------|-----|------------|-----|------------|-----|
| DIAGNOSTIC | rs2075650   | N  | rs4900200 | N  | rs13395562     | N   | rs17802296      | N   | rs5904473       | N   | rs644755 | N   | rs10501140 | N   | rs11950355 | N   |
| AD         | AA          | 38 | AA        | 8  | AA             | 63  | AA              | 12  | AA              | 102 | AA       | 22  | AA         | 21  | AA         | 98  |
| AD         | GA          | 46 | AG        | 29 | GA             | 63  | AG              | 68  | GA              | 40  | AG       | 73  | AG         | 70  | CA         | 82  |
| AD         | GG          | 9  | GG        | 56 | GG             | 18  | GG              | 113 | GG              | 51  | GG       | 98  | GG         | 102 | CC         | 13  |
| MCI        | AA          | 93 | AA        | 8  | AA             | 131 | AA              | 23  | AA              | 135 | AA       | 22  | AA         | 22  | AA         | 123 |
| MCI        | GA          | 70 | AG        | 79 | GA             | 149 | AG              | 83  | GA              | 42  | AG       | 125 | AG         | 123 | CA         | 97  |
| MCI        | GG          | 16 | GG        | 92 | GG             | 33  | GG              | 143 | GG              | 72  | GG       | 102 | GG         | 104 | CC         | 29  |
| NC         | AA          | 79 | AA        | 6  | AA             | 94  | AA              | 14  | AA              | 118 | AA       | 15  | AA         | 15  | AA         | 76  |
| NC         | GA          | 25 | AG        | 31 | GA             | 81  | AG              | 66  | GA              | 31  | AG       | 83  | AG         | 84  | CA         | 92  |
| NC         | GG          | 2  | GG        | 69 | GG             | 16  | GG              | 110 | GG              | 41  | GG       | 92  | GG         | 91  | CC         | 22  |

**A**



**B**



## Conclusões

## Conclusões

Após utilizar uma sequência de etapas associadas a diferentes ferramentas de bioinformáticas, foram encontradas variações previamente descritas no banco de dados de SNP do *National Center for Biotechnology Information* (dbSNP – NCBI), além de polimorfismos do tipo SNP e *INDELs* que surgiram como candidatos inéditos, para a manifestação dos TH. Estes achados levantam a possibilidade de que deleções de até 7 pares de bases estejam relacionadas à fisiopatologia dos TH, eventualmente em genes dos sistemas glutamatérgico e gabaérgico de neurotransmissor.

Com a evolução das abordagens de sequenciamento de nova geração elevaram-se o número de *INDELs* associados à doença genética humana. Assim, *SNPs* e *INDELs* podem ser utilizados como marcadores em busca de fatores genéticos que influenciem na fisiopatologia das doenças neuropsiquiátricas.

Vários *softwares* têm sido desenvolvidos para prever a relação entre uma variação genética e sua possível influência patogênica no indivíduo portador. Assim, a escolha do método computacional para que se possa ter um resultado *in silico* que melhor represente o acontecimento biológico, é parte inerente do sequenciamento do genoma humano e principalmente do mapeamento genético de doenças complexas, uma vez que essas doenças envolvem a regulação de muitos genes e possíveis interações entre eles e com o ambiente.

Os experimentos com dados de GWAS indicam que as ferramentas computacionais utilizadas são adequadas, para análise genética proposta. Os programas R e PLINK permitiram analisar os dados em tempo hábil e extrair informações já relatadas, forneceram dados inéditos para nosso trabalho, como no caso do SNP (rs5904473) localizado no gene *CCDC112* que até o presente trabalho não tinha sido associado a demências. Além disso, a nossa análise demonstra que características quantitativas podem revelar *insights* adicionais sobre o mecanismo que liga os níveis das *Aβ-42* no *CSF* e os níveis da *Aβ-40* no *CSF* e plasma sanguíneo e os genes candidatos apresentados como candidatos no envolvimento da fisiopatologia da DA e MCI.

Em suma, as dificuldades encontradas na identificação de polimorfismos genéticos de risco a doenças complexas mostram que ainda há lacunas a serem preenchidas que possibilitem a identificação de polimorfismos em larga escala. Quanto aos desafios computacionais, há uma ponte entre biologia e estatística que ainda precisa ser encontrada para que a biologia dê significado e embasamento aos resultados estatísticos e computacionais obtidos. No entanto, os resultados desse trabalho também mostram eficiência, rapidez e ampla aplicabilidade dos métodos computacionais na compreensão e interpretação de dados biológicos.

# Anexos

## Trabalhos Futuros

- Estudo de caso-controle e validação dos polimorfismos candidatos associados aos TH e a DA, na população brasileira;
- Como alternativa para análise em banco de dados, pretendemos combinar base de dados de *microarray* e *GWAS* para priorizar *SNPs* e explorar mais características das redes de interação gene-gene, que conduzam a associações relevantes com os fenótipos;
- Em estudos futuros, será também importante incluir as análises de epistasia entre os genes, para investigar associações conjuntas com os fenótipos;
- Investigar novas métricas de seleção de características, ou mesmo propor modificações nas métricas já existentes, para possíveis aplicações e consequente, comparações com os resultados aqui apresentados.

## Trabalhos Complementares

### Artigos completos publicados em periódicos

- Lima, S. G., ALBUQUERQUE, M. F. P. M., OLIVEIRA, J. R. M., Ayres, Constância F J, CUNHA, J. E. G., de Oliveira, Danyllo F, LEMOS, R. R., SOUZA, M. B. R., ODWALDO, S. Exaggerated blood pressure response as a risk factor for hypertension. Brazilian Journal of Medical and Biological Research (Impresso). , 2013.
- de Lima, Sandro G, de Albuquerque, Maria de F P M, de Oliveira, João R M, Ayres, Constância F J, da Cunha, José E G, de Oliveira, Danyllo F, de Lemos, Roberta R, de Souza, Manuela B R, e Silva, Odwaldo B. Exaggerated blood pressure response during exercise treadmill testing: functional and hemodynamic features, and risk factors. Hypertension Research. 2012.
- LEMOS, R. R., Lima, S. G., CUNHA, J. E. G., OLIVEIRA, D. F., AYRES, C. F. J., SOUZA, M. B. R., ALBUQUERQUE, M. F. P. M., OLIVEIRA, J. R. M. Revising the M235T polymorphism position for the AGT gene and reporting a modifying variant at the Brazilian population with potential cardiac and neural impact. Journal of Molecular Neuroscience. 2012.
- Queiroz, R.C., SOUZA, M. B. R., OLIVEIRA, J. R. M.  
A PILOT ANALYSIS FOR POPULATIONAL DISTRIBUTION OF THE 5-HTTLPR POLYMORPHISM COMPARED TO A GEOGRAPHIC DATABASE OF SUBJECTIVE PERCEPTION OF HAPPINESS. Neurobiologia (Recife. Impresso). 2011.

Trabalhos publicados em anais de eventos (resumo expandido)

- ARAUJO, G. S., COSTA FILHO, I. G., SOUZA, M. B. R., OLIVEIRA, J. R. M.  
An Experimental Application of Random Forest on ADNI Genotype Dataset Gilderlanio In:  
VII Simpósio Brasileiro de Bioinformática, 2012, Campo Grande.  
VII Simpósio Brasileiro de Bioinformática. , 2012.

Trabalhos publicados em anais de eventos (resumo)

- OLIVEIRA, D. F., Lima, S. G., CUNHA, J. E. G., LEMOS, R. R., SOUZA, M. B. R.,  
AYRES, C. F. J., ALBUQUERQUE, M. F. P. M., OLIVEIRA, J. R. M. Revisiting the  
M235T polymorphism position for the AGT gene and reposting a possible modifying  
variant at the Brazilian population In: 57° Congresso Brasileiro de Genética, 2011, Águas de  
Lindóia.  
57° Congresso Brasileiro de Genética. , 2011.
- Lima, S. G., ALBUQUERQUE, M. F. P. M., OLIVEIRA, J. R. M., AYRES, C. F. J.,  
CUNHA, J. E. G., LEMOS, R. R., OLIVEIRA, D. F., SOUZA, M. B. R. Ausência de  
associação entre o polimorfismo I/D da ECA e HAS In: 65° Congresso Brasileiro de  
Cardiologia, 2010, Belo Horizonte.  
Arquivos Brasileiros de Cardiologia. , 2010. v.95. p.91 - 91
- Lima, S. G., ALBUQUERQUE, M. F. P. M., OLIVEIRA, J. R. M., AYRES, C. F. J.,  
CUNHA, J. E. G., LEMOS, R. R., OLIVEIRA, D. F., SOUZA, M. B. R. Elevação  
exagerada da pressão arterial no TE: o papel do polimorfismo I/D da ECA In: 65° Congresso  
Brasileiro de Cardiologia, 2010, Belo Horizonte.  
Arquivos Brasileiros de Cardiologia. , 2010. v.95. p.47 – 47.