
**Universidade Federal de Pernambuco – UFPE**

**Centro de Ciências Biológicas – CCB**

**Programa de Pós-Graduação em Ciências Biológicas**

**Doutorado em Ciências Biológicas**

**ROBERTA RODRIGUES DE LEMOS**

# Identificação das bases genéticas nas Calcificações Idiopáticas em Gânglios da Base do cérebro (IBGC) e *screening* de variações candidatas na Doença de Alzheimer

Orientador: Prof. Dr. João Ricardo Mendes de Oliveira

RECIFE 2012

**ROBERTA RODRIGUES DE LEMOS**

# Identificação das bases genéticas nas Calcificações Idiopáticas em Gânglios da Base do cérebro (IBGC) e *screening* de variações candidatas na Doença de Alzheimer

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

Orientador: Prof. Dr. João Ricardo Mendes de Oliveira

RECIFE 2012

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

L555i Lemos, Roberta Rodrigues de

Identificação das bases genéticas nas calcificações idiopáticas em gânglios da base do cérebro (IBGC) e *screening* de variações candidatas na Doença de Alzheimer / Roberta Rodrigues de Lemos. – Recife: O Autor, 2012.

93 folhas: il., fig., tab.
Orientador: João Ricardo Mendes de Oliveira
**Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Ciências Biológicas. Pós-graduação em Ciências Biológicas, 2012.**
Inclui referências e anexos

**1. Alzheimer, Doença de 2. Genética 3. Bioinformática I. Oliveira, João Ricardo Mendes de (orientador) II. Título.**

*616.831 CDD (22.ed.) UFPE/CCB-2013-002*

**ROBERTA RODRIGUES DE LEMOS**

# Identificação das bases genéticas nas Calcificações Idiopáticas em Gânglios da Base do cérebro (IBGC) e *screening* de variações candidatas na Doença de Alzheimer

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

Banca Examinadora

_____

Prof. Dr. João Ricardo Mendes de Oliveira (Orientador)
Universidade Federal de Pernambuco (UFPE)

_____

Dra. María Jesús Sobrido Gomes
Fundación Pública Galega de Medicina Xenómica (Espanha)

_____

Dra. Gabriela Ferraz
Instituto de Medicina Integral Prof. Fernando Figueira (IMIP)

_____

Dr. Carlos Henrique Madeiros Castelletti
Instituto Agronômico de Pernambuco (IPA)

_____

Profa. Dra. Danyelly Bruneska Godim Martins
Universidade Federal de Pernambuco (UFPE)

RECIFE 2012

“Any sufficiently advanced technology is
indistinguishable from magic.”
Arthur C. Clarke

À minha mãe Janette, um exímio de
perseverança e amor, a toda minha família e amigos,
Dedico.

## AGRADECIMENTOS

"A única forma de chegar ao impossível é acreditar que é possível" (Alice no País das Maravilhas), principalmente com incentivo daqueles que te amam e que apostam no teu sonho, nesse sentido, agradeço:

Ao Deus de infinito amor, por me indicar o caminho certo e proporcionar encontros espetaculares, com pessoas maravilhosas, durante essa trajetória.

À minha mãe, Janette, exemplo de determinação, disciplina e amor. Obrigada mãe, por tanta dedicação gratuita, por tanto incentivo, por mostrar sempre o lado bom das situações e principalmente por acreditar.

Aos meus eternos avós: Ivanise e Epaminondas Campelo (em memória), exemplos de puro amor.

A minha Família, laço indissolúvel: irmão, pai, cunhada, sobrinha, tias, tios, primos, primas, a todos minha gratidão pela grande torcida e compreensão.

A meu grande amor, amigo e companheiro Luiz Roberto Gitirana (Beto), um ser iluminado que entrou na minha vida como um presente divino, agradeço por sua paciência, dedicação, compreensão e amor.

A meu orientador e amigo João Ricardo, por acreditar e investir na minha formação profissional e, sobretudo pessoal. Muito obrigada pelo seu contínuo exemplo de paciência, sabedoria, alegria e perseverança.

La Dra María Jesús Sobrido (Susi), un ejemplo científico, te lo agradezco por sú dedicación y apoyo, por la amistad sincera.

Aos meus queridos amigos, estes compartilharam comigo, suas sabedorias, suas grandezas, suas alegrias. Minhas amigas da graduação: Carla Maciel e Suzy Quirino. Meus amigos do intercâmbio: Fabiana, Alessandra, Laís de Paula, Laís Galdino, Ana, Lilian, Sebastián, Pilar, Andrés, María, Lorena, Luiza, Isadora, Juliana. A equipe do "Sharing Galícia" que proporcionou passeios belíssimos e a Profª Rita Fabiana, imprescindível para a rápida aprendizagem da língua Espanhola.

Aos meus amigos do Grupo Psiquiatria Biológica/UFPE: Eriton, Lylyan, Manuela, Regina, Danyllo, Joana, Samy, Matheus e outros. Em particular a Eriton, por todo o suporte e paciência, a Regina e a Lylyan por todo carinho e dedicação dado a nossa amizade, a todos vocês obrigada pelos sorrisos, pelas doces palavras e pelos ensinamentos.

A Profª Dra Mayana Zatz (USP), Profª Dra Maria de Mascena (UFRPE), Profª Dra Maria Tereza Cartaxo (UPE) e Profª Dra. Flavia de Paula (UFES) por todo o incentivo, colaboração, apoio técnico e financeiro.

Aos amigos do Laboratório de Imunopatologia Keizo Asami (LIKA): Profº Dr. José Luiz, Profª Dra. Cíntia Rocha, Profª Dra. Danielle Bruneska, Dr. Henrique Castelletti, Profª Dra. Rosalie Belian, Eliete e Claúdio pelos ensinamentos, suporte e, sobretudo amizade.

Aos programas de Pós-Graduação: do Centro de Ciências Biológicas – CCB/ UFPE e do LIKA/UFPE, pelo apoio e suporte financeiro.

Aos programas de fomento: CAPES, FACEPE, PROPESQ-UFPE, Fundação Carolina pelo apoio financeiro importantíssimo para pesquisa Brasileira.

**Resumo**

A Degeneração Neuronal (DN) é um processo complexo e multifatorial, presente em várias desordens neuropsiquiátricas, como a Doença de Fahr (DF) e a Doença de Alzheimer (DA), entre outras. Este trabalho teve como objetivo, estudar fatores de risco genéticos, em dois modelos de doenças neurodegenerativas, através de sequencimentos e métodos de bioinformática. O *screening* de genes e variações candidatas foi realizado através de Reação em Cadeia da Polimerase (PCR), seguido de Sequenciamento Automático Direto do DNA. Para a DF, também conhecida como Calcificação Idiopática dos Gânglios da Base (IBGC) dois genes foram investigados, o gene MGEA6/CTAGE lócus IBGC1 e o gene SLC20A2 lócus IBGC3. As variações candidatas para DA foram selecionadas a partir de um *pipeline* de bioinformática, onde informações genéticas de diferentes bancos de dados foram integradas, sendo os principais genes estudados: COX-2, IL1a e CD83. Os resultados das genotipagens não identificaram, em ambos os grupos de controles e afetados, a variação Pro521Ala (rs36060072) / MEGEA6 lócus IBGC1. Essas genotipagens contribuíram para a diminuição da freqüência alélica mínima dessa variante, demonstrando o quanto ela é rara e ausente nessa amostra da população Brasileira. Para o segundo lócus analisado, IBGC3 gene SLC20A2, os achados tem mostrado novas variações em diferentes exons e estes estão associadas a mais de 40% das famílias com IBGC analisadas. No entanto, algumas famílias recrutadas foram excluídas para esses dois loci, sugerindo adicional heterogeneidade genética. Paralelamente, os resultados do ensaio experimental do *screening* das variações candidatas para DA, só foi permito em parte, pois apenas variações do tipo SNPs: "rs2745557"_COX2 (exon 1) e "rs17561"_IL1A (exon5), foram detectadas, sendo a grande maioria das variações preditas, INDELs (1 a 10pb), não encontradas. O desafio atual é aperfeiçoar os métodos de sequenciamentos, principalmente através do uso novas plataformas de sequenciamentos. Há uma grande expectativa de que um melhor entendimento, das bases biológicas destes dois distúrbios investigados, possa contribuir para a compreensão amplamente de outras doenças neuropsiquiátricas.

**Palavras chave:** Calcificações idiopáticas dos Gânglios da Base, Doença de Alzheimer, Bioinformática, INDELs, gene SLC20A2.

**Abstract**

Neuronal Degeneration (DN) is a complex and multifactorial, present in several neuropsychiatric disorders such as Fahr's Disease (FD) and Alzheimer's Disease (AD), among others. This work aimed to study genetic risk factors, in two models of neurodegenerative diseases through sequencing and bioinformatics resources. The screening of candidate genes and variants was performed by Polymerase Chain Reaction (PCR), followed by Auto Direct Sequencing of DNA. For FD, also known as Idiopathic Basal Ganglia Calcification (IBGC) two genes were investigated, the MGEA6/CTAGE gene IBGC1 locus and SLC20A2 gene locus IBGC3. Variations for AD candidates were selected from a bioinformatics pipeline, where genetic information from different databases was integrated, the main genes studied: COX-2, CD83 and IL1a. The results of the genotyping not identified in both groups of controls and affected the variation Pro521Ala (rs36060072) / MEGEA6 IBGC1 locus. These genotyping contributed to reducing the Minimum Allele Frequency (MAF) of this variant, demonstrating how it is rare or absent in this sample of the Brazilian population. For the second IBGC3 locus SLC20A2 gene, new findings have shown variations in different exons and these are associated with more than 40% of families with IBGC analyzed. However, some families were excluded recruited to these loci, suggesting further genetic heterogeneity. In parallel, the assay results of variations screening candidates for AD, allow only variations in type SNPs "rs2745557" _COX2 (exon 1) and "rs17561" _IL1A (exon5) were detected, and the vast majority of predicted variations, INDELs (1 to 10pb) not found. The current challenge is to improve methods of sequencing, primarily through the use of new sequencing platforms. There is a great expectation that a better understanding of the biological basis of these two disorders investigated, may largely contribute to the understanding of other neuropsychiatric diseases.

**Key words:** Idiopathic Basal Ganglia Calcification, Alzheimer Disease, Bioinformatics, INDELs, SLC20A2 gene.

**Lista de Figuras da Revisão Bibliográfica**

**Capítulo 1.** Revisão Bibliografia

**Lista de Tabelas da Revisão Bibliográfica**

**Capítulo 1.** Revisão Bibliografia

**Lista de Figuras de cada capítulo**

**Lista de Tabelas de cada capítulo**

**Lista de Abreviações**

ABRAZ – Associação Brasileira de Alzheimer

DNA - Deoxyribonucleic Acid - Ácido desoxirribonucléico

DF - Doença de Fahr

DA - Doença de Alzheimer

DP - Doença de Parkinson

DFT - Degeneração Fronto Temporal

DN - Degeneração Neuronal

dbSNP - data base Single Nucleotide Polymorphism – banco de dados de Polimorfismos de Base Única.

ELA - Esclerose Amiotrófica Lateral

EOAD - Early Onset Alzheimer´s Disease - Doença de Alzheimer de Início Precoce

EST - Expressed Sequence Tags - Sequencias de Etiquetas Expressas

IBGC - Idiopathic Basal Ganglia Calcification - Calcificação Idiopática dos Gânglios da Base

INDELs - Insertion/Deletion – Inserções /Deleções

LOAD - Load Onset Alzheimer´s Disease - Doença de Alzheimer de Início Tardio

MAF - Minor Allele Frequency - Freqüência Alélica Mínima

NCBI- National Center for Biotechnology Information- Centro Nacional de Informação Biotecnologica

NGS - Next Generation Sequencing - Sequenciamento de Nova Geração

PCR - Polymerase Chain Reaction - Reação em Cadeia da Polimerase

TC - Tomografias Computadorizadas

WHO - World Health Organization - Organização Mundial de Saúde

**Sumário**

## 1.0 Introdução

Há um grande interesse em se desvendar as bases genéticas dos transtornos neuropsiquiátricos, com manifestações comportamentais e motoras.

As Calcificações Idiopáticas dos Gânglios da Base do cérebro (IBGC), também conhecida como "Doença de Fahr" - DF é uma doença neurodegenerativa, sem associação com outros distúrbios metabólicos e que apresenta diversos sintomas associados, tais como: enxaqueca, parkinsonismo, psicose, demência e alterações de humor.

Até pouco tempo IBGC era considerada uma doença rara, mas a disponibilização e a redução dos custos dos exames de neuroimagem (tomografias, ressonâncias) têm permitido um relato cada vez maior de pacientes.

Esta investigação pretende dar continuidade ao avanço da compreensão das bases genéticas e moleculares em IBGC, bem como aprofundar o entendimento das relações genótipos e fenótipos.

A Doença de Alzheimer – DA, foi outro modelo de doença neurodegenerativa explorado nesse trabalho. A DA é caracterizada como um distúrbio progressivo da memória, sendo a causa mais comum de demência no mundo.

Para a DA, foi proposto validar um *pipeline* de bioinformática, gerado por nosso grupo. Nesse *pipeline* as informações de bancos de dados genéticos foram integradas, a fim de gerar um painel de variações potencialmente candidatas para DA.

Integrar abordagens computacionais as técnicas de biologia molecular para explorar os transtornos neuropsiquiátricos, tem se tornado cada vez mais relevante, uma vez que essas doenças são um dos principais problemas de saúde pública em todo o mundo, além de apresentar caráter crônico e progressivo, associados à alta morbidade e mortalidade.

### 1.1. Justificativas

Dentre as doenças neurodegenerativas, a DA está atualmente entre os principais problemas de saúde pública em todo o mundo e com previsão de aumento devido à elevação da expectativa de vida.

A importância do estudo da DF se dá pelas evidências atuais que sugerem que os depósitos de cálcio em gânglios da base do cérebro estejam presentes em cerca de 1% das Tomografias Computadorizadas- TCs e que estejam relacionados com sintomas neuropsiquiátricos comuns, tais como transtornos do humor e psicose.

Investigar fatores de risco genético, através do uso de técnicas de sequenciamento, tem se tornado uma abordagem massiva, principalmente pelo uso das plataformas ditas *Next Generation Sequencing* - NGS, onde os inúmeros dados gerados são processados por ferramentas de bioinformática.

As análises desenvolvidas neste projeto poderão auxiliar, tanto na definição de métodos diagnósticos mais rápidos e eficazes, como na melhor compreensão da fisiopatologia da DA e da DF, abrindo novas perspectivas para o tratamento baseado em evidências genéticas.

## 1.2 Objetivos

### 1.2.1 Geral

Identificar as bases genéticas da DF e realizar o *screening* de mutações candidatas na DA, através de genotipagem em larga escala, com o uso de técnicas de Sequenciamento e Bioinformática

### 1.2.2 Específicos

- Triar em amostras de DNA humano da população Brasileira, variações candidatas preditas por técnicas de Bioinformática;
- Validar experimentalmente o *pipeline* de Bioinformática, desenvolvido a partir de bancos de dados genéticos;
- Triar em amostras de DNA humano da população Brasileira a mutação P521A, presente no lócus IBGC1 associado à DF;
- Investigar as bases genéticas da DF.

# Capítulo 1

# Revisão Bibliográfica

## 1. Neurodegeneração e Genética

A degeneração neuronal (DN) é um fenômeno complexo com etiologia multifatorial e presente em várias doenças neuropsiquiátricas, tais como: Doença de Alzheimer (DA), Doença de Parkinson (DP), Demência Fronto Temporal (DFT), Doença de Fahr (DF), Esclerose Amiotrófica Lateral (ELA), entre outras (GOLDEMAN, 2012; SÁNCHEZ-PÉREZ *et al*., 2012; PASCHALI et al., 2009).

O fenômeno da DN é evidenciado por sinais e/ou sintomas progressivos como as demências, com um amplo espectro de início, sendo mais freqüente na idade adulta, provocando lesões irreversíveis que já foram bem caracterizadas em análises anatomopatológicas e radiológicas (WINKHOFER *et al*., 2008, LEMOS *et al*., 2011; CRUTS *et al*., 2012).

Aproximadamente 35,6 milhões de pessoas no mundo sofrem com algum tipo de demência, e as estimativas apontam que em 2030 serão 65,7 milhões e em 2050 serão 115,4 milhões portadores (WHO, 2012). A DA é a causa mais comum de demência, e possivelmente contribui para até 70% dos casos. No Brasil o número estimado de pacientes com DA é de 1 milhão e 200 mil afetados (ABRAZ, 2011).

A contribuição genética para as doenças neurodegenerativas resulta em uma complexa combinação de múltiplos fatores de riscos e de proteção genéticos, que são orquestrados com fatores ambientais, influenciando no risco do indivíduo desenvolver a doença em um determinado estágio da vida (CRUTS *et al*., 2012).

Estudos genéticos de associação caso-controle para DA apontam mais de 500 genes associados a essa doença, em diversas etnias, esses dados estão catalogado em um banco de dados público, chamado Alzgene, disponível pelo site (www.alzforum.org).

O acervo de dados sobre fatores de risco genético também se destaca para outras doenças, tais como: DP (www.pdgene.org); DFT (www.molgen.ua.ac.be/ADMutations) e ELA (www.alsod.iop.kcl.ac.uk).

No contexto, investigação de doenças neurodegenerativas, nosso grupo de pesquisa, vem contribuindo a mais de 10 anos, do ponto de vista: clínico, molecular, bioinformático, e mais recentemente através de análises de neuroimagens. Para alguns temas, como a DA e a DF, os estudos procuram esclarecer o cenário genotípico e fenotípico (OLIVEIRA *et al*., 2007, 2009a, 2009b, 2009c, 2009d; SOUZA *et al*., 2010; LEMOS *et al*., 2011, 2012a; WANG *et al*., 2012).

## 2. IBGC (Calcificação Idiopática dos Gânglios da Base)

2.1 Características gerais

Historicamente o termo "Doença de Fahr" se tornou mais popular do que Calcificação Idiopática dos Gânglios da Base (IBGC), devido à descrição de Karl Theodor Fahr (1877-1945). Em 1930, Fahr descreveu um paciente que apresentava sintomas como: demência, febre, imobilidade sem paralisia, tosse e úlcera decúbito, com falecimento logo após o terceiro dia de admissão, a autopsia revelou núcleos de calcificações nos gânglios da base (MANYAM, 2005).

Fahr, todavia não foi o primeiro a descrever achados de depósitos de cálcio, principal mineral dessas calcificações, sendo que outros autores já haviam relatado casos sobre esse tema; Delacour em 1850, Bamberger em 1855 e Later em 1902 (MANYAM, 2005).

Os gânglios da base são caracterizados por estruturas subcorticais compostas principalmente de corpos celulares neuronais. Estes estão envolvidos com diversas funções cognitivas, tais como: coordenação, memória e pensamento (MANYAM, 2005).

IBGC se refere à formação de depósito de minerais, principalmente na forma de fosfato de cálcio, nos gânglios da base do cérebro (Figura 1) e ocasionalmente em outras estruturas como: cerebelo, substância branca e tálamo (OLIVEIRA *et al*., 2004; MANYAM, 2005).

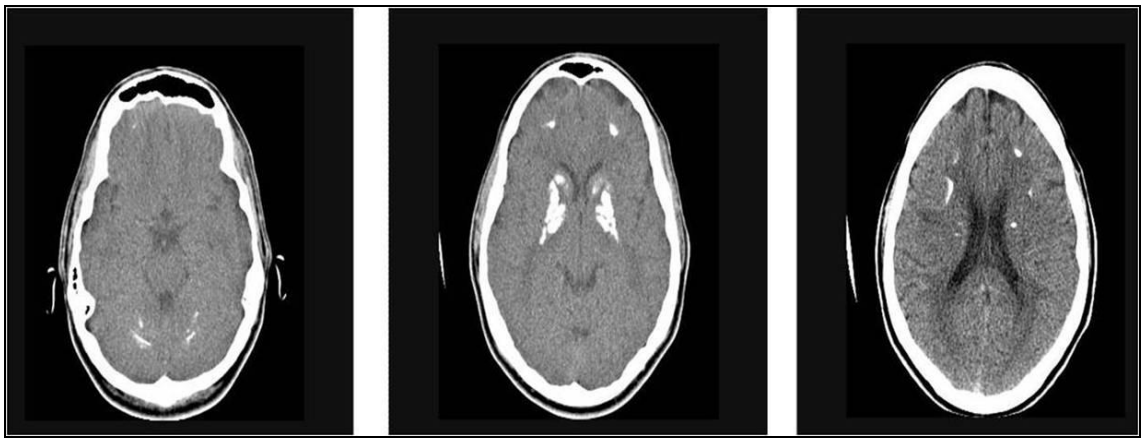


Figura 1. Imagens 2D de Tomografias Computadorizadas de um paciente com IBGC. As calcificações podem ser observadas em cerebelo, gânglios da base e matéria branca (Imagens do artigo de OLIVEIRA *et al*., 2009d).

A idade de início varia entre 30 e 60 anos, mas já foram descritas crianças com IBGC (GESCHWIND *et al*., 1999; SOBRIDO *et al*., 2007). Vários estudos mostram que há uma diferença na prevalência dos achados de calcificações nos gânglios da base,

variando de 0.3% até 12%, em exames de neuroimagem (MANYAM, 2005; FUJITA *et al*., 2003).

Um estudo sobre a prevalência de calcificações nos gânglios da base, na cidade do Recife/PE, aponta que os achados de calcificações estão em torno de 2,42%, em 1898 exames de TCs. A alta taxa de calcificação se destaca para a faixa etária acima dos 60 anos, sugerindo relação com o processo de envelhecimento e morte neuronal (BARRO e SILVA, 2008).

Exames imuno-histoquímicos de calcificações intracranianas em pacientes com DA, DP, DFT e Degeneração Neurofibrilar Difusa com Calcificações, mostram depósitos de proteínas de matriz óssea não colagenosa, tais como; osteocalcina, osteopontina, osteonectina e sialoproteina (FUJITA *et al*., 2003).

Por isso, há uma grande expectativa de que o melhor entendimento das bases biológicas da IBGC possa contribuir de maneira ampla na compreensão de outros quadros afins, como a esquizofrenia, o parkinsonismo, as demências e os transtornos do humor, considerando as sobreposições clínicas e histopatológicas entre a IBGC e estas outras doenças neuropsiquiátricas.

## 2.2 Aspectos Clínicos

A IBGC deve ser vista como uma condição neuropsiquiátrica, onde o indivíduo pode simultaneamente apresentar sintomas neurológicos e psiquiátricos.

A manifestação clínica da IBGC é heterogênea, sendo caracterizada por parkinsonismo, distonia e sintomas neuropsiquiátricos como; psicose, quadros demenciais e transtornos de humor, onde o indivíduo apresenta taxas normais endocrinológicas e geralmente um padrão de herança autossômico dominante (MANYAM, 2005).

Sintomas de tontura, vertigem e dores de cabeça também estão relacionados a este transtorno que apresenta amplo espectro de manifestações clínicas entre diferentes famílias e até mesmo dentro de uma mesma genealogia (GESCHWIND *et al*., 1999; SOBRIDO *et al.*, 2007).

De acordo com Bala Manyam, os critérios de inclusão para diagnostico de IBGC são: (1) evidências radiológicas de calcificações bilaterais, quase simétricas, seguidas de uma ou mais áreas; gânglios basais, núcleos denteado, tálamo, e matéria branca cerebral; (2) crescimento e desenvolvimento do indivíduo normal durante a infância; (3) ausência de desordens paratireóides; (4) detalhes de informações sobre o *pedigree*

familiar; (5) documentação detalhada sobre a exploração histórica e clínica (OLIVEIRA, 2011).

### 2.2.1 Penetrância Incompleta e Resiliência Cerebral

Na IBGC, a heterogeneidade fenotípica é uma característica marcante, dada pela observação dos achados de calcificações e investigações clínicas (OLIVEIRA *et al*., 2009d).

Na clínica, são observados indivíduos assintomáticos ou com suaves sintomas, que possuem diferentes níveis de calcificações (Figura 2 A, B, E), os estudos sugerem que é preciso levar em consideração dois mecanismos de penetrância; um para a formação da calcificação e outro para os primeiros sinais da manifestação clínica (OLIVEIRA *et al*., 2009d; OLIVEIRA, 2011).

Particularmente, essas famílias (Figura 2 A, B, E) apresentam adicionais calcificações no cerebelo e curiosamente, um mecanismo de resiliência cerebral modulado pelo cerebelo, foi sugerido a partir de um estudo de neuroimagem de pacientes com Distonia (ARGYLEAN *et al*., 2009).

Argylean *et al*. (2009) investigaram o circuito cerebelo-tálamo-cortex (cbTC), através de técnicas de neuroimagem, em indivíduos com Distonia sendo, sintomáticos e assintomáticos, ambos grupos portadores de mutações, e comparou com indivíduos controles. Diferenças entre os grupos em relação ao circuito foram observadas. Os resultados apontam que as alterações neste circuito poderiam estar "desligando" alguns pontos de saída da transmissão normal, desta forma silenciando as manifestações clínicas nos indivíduos assintomáticos (ARGYLEAN *et al*., 2009).

Há uma complexa relação entre: as variações no genoma humano, os limites dos efeitos da expressão de doenças neuropsiquiátricas e a manutenção da resiliência cerebral. Neste caso, resiliência, é a capacidade do cérebro permanecer sem manifestar sinais de diminuição da função, apesar dos prejuízos de origem genética (KEMPTOM *et al*., 2009).

Resiliência e Penetrância incompleta são mecanismos sugeridos e que se completam na fisiopatologia da IBGC (OLIVEIRA, 2011).

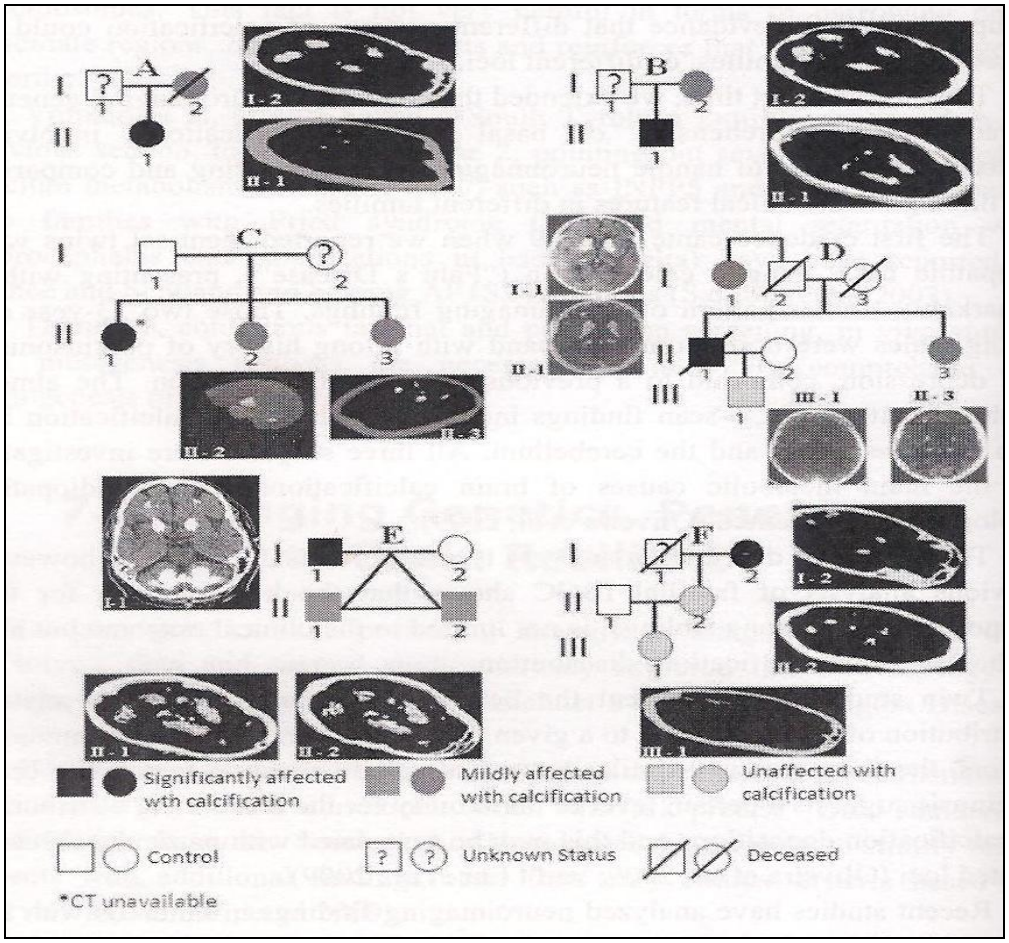


Figura 2. Imagens de Famílias Brasileiras com IBGC, essas imagens são reconstruções tridimensionais de fatias de neuroimagens (Imagem do livro Oliveira, 2011).

2.2.2 Estudos de Neuroimagem Funcional

Interpretar a expressão de fenótipo clínico apenas como medida de penetrância de um gene ou genes mutados, pode ser incompreensível para os casos de predisposição genética e indivíduos assintomáticos, onde há notáveis disfunções observadas através de estudos de neuroimagem funcional (OLIVEIRA, 2011).

Esses estudos podem detectar previamente disfunção cortical em pacientes com IBGC, sendo eles sintomáticos e assintomáticos, através da avaliação do fluxo sanguíneo nos gânglios da base (PASCHALI *et al*., 2009; SAIKI *et al*., 2007).

A via dopaminérgica e o fluxo sanguíneo cerebral regional foram avaliados em uma paciente de 56 anos de idade com IBGC que, apresentava distúrbios de movimentos, prejuízo da memória e da fala. Os marcadores usados foram o Transportador de Dopamina (DAT) combinado com baixa dose de raio-x de tomografia computadorizada de transmissão (SPECT/TC) e $99^{m}$ Tc-D,L hexamethylpropylene amine oxime ($99^{m}$ Tc-D,L-HMPAO). Os resultados revelaram hipoperfusão nas regiões

calcificadas dos gânglios da base, demonstrando correlação com a condição clínica da paciente (PASCHALI *et al*., 2009).

No entanto, um estudo de neuroimagem funcional demonstrou perfusão normal em uma paciente de 82 anos de idade com IBGC sendo, assintomática e com limitada calcificação no núcleo denteado e gânglios da base. Nesse caso a presença de calcinose não pode ser critério para danos neurológicos, mas o grau de calcinose e a redução da perfusão pode ser crítico para separar sintomáticos dos assintomáticos com IBGC (SAIKI *et al*., 2007).

## 2.3 Aspectos Genéticos

A maioria dos casos de IBGC apresenta um padrão de herança autossômico dominante (SOBRIDO *et al*., 2007). O primeiro lócus associado à IBGC foi localizado no braço longo do cromossomo 14 - IBGC1(GESCHWIND *et al*., 1999).

No entanto, outras famílias com IBGC foram excluídas da região candidata no cromossomo 14 indicando uma possível heterogeneidade genética (OLIVEIRA *et al*., 2004).

O *screening* de possíveis mutações no lócus IBGC1 identificou uma variação em heterozigose (rs36060072), potencialmente patogênica no gene MGEA6/CTAGE5, presente em todos afetados de uma grande genealogia Americana e ausente nos controles (GESCHWIND *et al*., 1999; OLIVEIRA *et al*., 2007).

A continuidade das pesquisas apontou mais outros loci candidatos à IBGC nos cromossomos 2 (IBGC2), 8 (IBGC3), 7 e 9, (Tabela 1) demonstrando heterogeneidade genética (GESCHWIND *et al*., 1999; OLIVEIRA *et al*., 2003, 2007, 2009a, 2009b, 2009c; VOLPATO *et al*., 2008, 2009; DAI *et al*., 2010, WANG *et al*., 2012).

Tabela 1. Heterogeneidade genética observada na IBGC. Principais loci associados a essa doença.

| **Loci** | **Autores** |
|---|---|
| IBGC1 / Cromossomo 14 | GESCHWIND *et al*., 1999; OLIVEIRA *et al*.,2007 |
| IBGC2/ Cromossomo 2 | VOLPATO *et al*., 2008; 2009. |
| IBGC3/ Cromossomo 8 | DAI *et al.*, 2010; WANG *et al*., 2012 |

### 2.3.1 Mutações no Gene SLC20A2 (PiT2)

Wang *et al*. (2012) identificaram sete novas mutações em pacientes com IBGC nas famílias da China, Espanha e Brasil, essas mutações estão no gene SLC20A2 no

lócus IBGC3, esse gene codifica o transportador de fosfato 2 chamado de PiT2 (Figura 3).

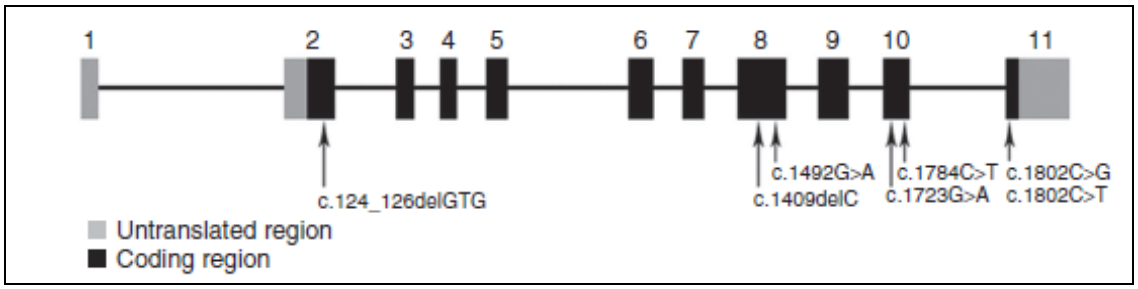



Figura 3. Estrutura do gene SLC20A2. As setas indicam as mutações detectadas nas regiões exônicas. (Imagens do Artigo de WANG *et al*., 2012).

O gene SLC20A2 e seu homólogo SLC20A1 (PiT1) fazem parte da família tipo III de transportadores de fosfato dependente de sódio, esses dois genes são constitutivamente expressos em vários tecidos, inclusive no cérebro (LAGRUE *et al*.,2010).

Estudos funcionais do padrão de expressão dos genes SLC20A2 e SLC20A1 no rim demonstram uma intrínseca relação desses genes com a concentração de fosfato inorgânico (Pi) sendo, PiT1 e PiT2 responsáveis pelo controle da receptação do Pi sob diferente condições de $p^H$ dieta de Pi (KAVANAUGH & KABAT, 1996; VILLA-BELLOSTA *et al*., 2009).

Outros estudos funcionais demonstram a relação do PiT1 e não do PiT2 com calcificação das células do músculo liso vascular, sendo Pi o principal indutor da calcificação vascular (VILLA-BELLOSTA *et al*., 2007; GIACHELLI *et al*., 2009).

Evidências indicam que o aumento da concentração de Pi no soro, pode induzir calcificação das paredes dos vasos sanguíneos, essa associação foi relatada em processos cardiovasculares provocando morte de pacientes com doenças renais crônicas, bem como na população em geral (KENDRICK *et al*., 2011).

Os estudos funcionais feitos por Wang *et al*. (2012) em oócitos de *Xenupus leavis* sugerem que as mutações no gene SLC20A2, possam estar associadas com o fenômeno da haploisuficiência, resultando no aumento da concentração do Pi, e essa elevada concentração poderia induzir calcificação mediada por PiT1, uma vez que o PiT1 está mais relacionado com o processo de calcificação vascular (WANG *et al*., 2012). No entanto, mais estudos ainda são necessários para elucidar os mecanismos moleculares da IBGC.

A partir das evidências de Wang *et al*. (2012), foi dado início ao recrutamento de pacientes com IBGC e sequenciamento dos genes SLC20A2 e SLC20A1, os resultados dessas investigações serão discutidos nos capítulos 4 e 5 dessa tese.

## 3.0 Doença de Alzheimer

### 3.1 Caracterização da Doença de Alzheimer –DA.

DA é classificada como "*polyproteinphaties*" onde múltiplas proteínas assumem conformação potencialmente patogênica e acumula-se isoladamente ou em conjunto no cérebro (HUANG & MUCKE, 2012; YANKNER, 2008).

A forma precoce (*Early Onset Alzheimer's Disease* - EOAD) se inicia antes dos 65 anos, tem herança autossômica dominante e é identificada em grupos familiares específicos, correspondendo a aproximadamente 1-2% dos casos. Até o momento, várias mutações foram verificadas em três genes, sendo eles: proteína precursora amilóide (APP), presenilina-1 (PSEN-1) e presenilina-2 (PSEN-2), localizados nos cromossomos 21, 14 e 1 respectivamente (DUBOIS *et al*., 2007; PASTORINO *et al*., 2006; HEESE *et al*., 2006, GUERREIRO *et al*., 2010).

A APP gera o peptídeo β-amilóide que sofre clivagens após sua síntese, quando clivado pela α-secretase produz um fragmento solúvel, porém quando clivado pela γ-secretase gera o fragmento ($A\beta_{40-42}$) insolúvel e de fácil agregação, estudos indicam que mutações nas presenilinas também as torne semelhantes a enzima γ-secretase, provocando a geração do fragmento $A\beta_{40-42}$ (CHAU *et al*., 2012).

Dados recentes têm apontando mais de 170 mutações presentes no gene PSEN-1, esses dados e de outros genes ligados a forma precoce da DA, estão catalogados no (http://www.molgen.ua.ac.be/ADMutations), mutações nos genes APP, PSEN-1 e PSEN-2 correspondem a 30-50% dos casos autossômicos dominantes (DUBOIS *et al*., 2007; PASTORINO *et al*., 2006; HEESE *et al*., 2006, CHAU *et al*., 2012).

A forma tardia (*Late Onset Alzheimer's Disease* – LOAD) corresponde à maioria dos relatos em DA, ocorre após os 65 anos em mais de 13% dos indivíduos e tem o risco aumentado para 30-50% em indivíduos acima de 80 anos de idade, tendo como modelo a herança multifatorial com contribuição de fatores genéticos e ambientais. O principal gene relacionado a essa forma é a isoforma E4 da Apolipoproteína E (APOE4) correspondendo a 20% dos casos, sendo localizado no cromossomo 19 (GUERREIRO *et al*., 2010).

APOE é uma proteína multifuncional que desempenha funções no metabolismo lipídico e neurobiológico, essa proteína possui três isoformas (APOE2, APOE3 e APOE4) que tem papéis diferentes sobre o metabolismo lipídico e a homeostase neuronal (HUANG, 2006, HUANG & MUCKE, 2012).

O alelo APOE4 é o mais importante fator de risco para LOAD, modificando a evolução clínica do quadro demencial a resposta a agentes anti-colinesterásicos e modulando também modificações anatômicas e metabólicas detectáveis através de exames de neuroimagem (SMALL *et al*., 2000; OLIVEIRA *et al*., 1999, Alzheimer Association, 2009).

Estudos demonstram que APOE3 estimula o crescimento neurítico e a isoforma E4 inibe, além de causar prejuízos nas funções mitocondriais e estar associado com disfunções dos astrocitos (HUANG, 2006).

As isoformas E3 e E4 podem, *in vitro*, formar complexos estáveis com o fragmento peptídico Aβ, sendo que a isoforma E4 processa esse complexo mais rapidamente e eficientemente (KIN *et al*., 2009).

Curiosamente, estudos com modelos de ratos, demonstram que na ausência do acúmulo do fragmento Aβ, a presença da isoforma E4 pode ter efeitos prejudicais sobre as atividades neuronais e comportamentais (LI *et al*., 2009).

No entanto, essa variante genética, APOE4, está longe de explicar todo o processo neurodegenerativo e vários outros polimorfismos têm sido associados com essa devastadora doença que se manifesta com déficit de memória e alterações de comportamento.

Outro achado macromolecular associado à DA são os emaranhados neurofibrilares, que são constituintes intracelulares e se formam pela hiperfosforilação da proteína Tau (DEHMELT & HALPAIN, 2004).

Interessantemente, mutações no gene MAPT que codifica a proteína Tau resulta em doenças como a DFT, mas não DA. Na DA, o fragmento Aβ e a isoforma APOE4 atuam na forma selvagem da proteína Tau, alterando sua função e estrutura, através de modificações pós-traducionais (Figura 4), todavia o exato mecanismo que a Tau desenvolve nas doenças DFT e DA ainda precisa ser determinado (MORRIS *et al*., 2011; HUANG & MUCKE, 2012).

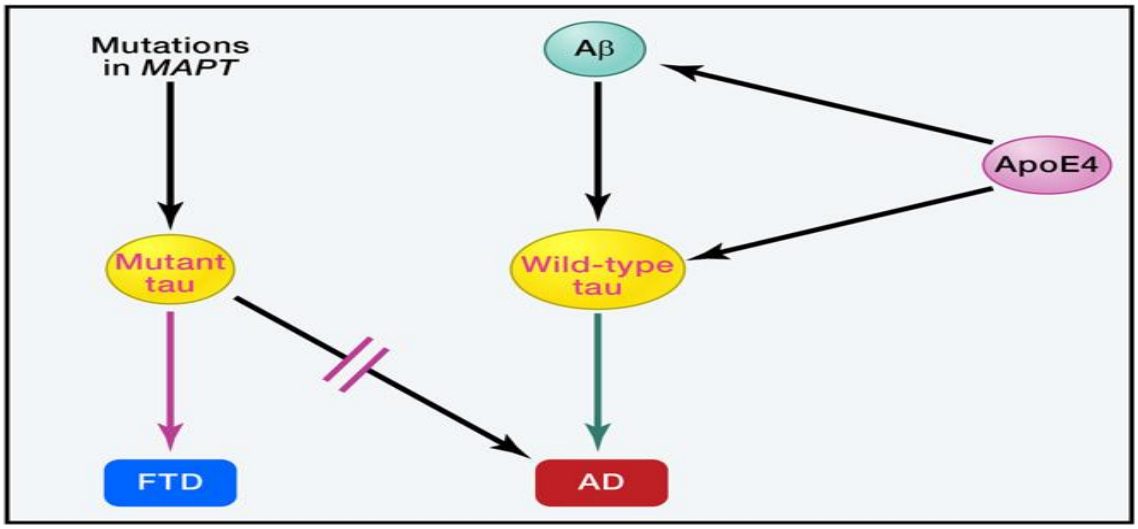



Figura 4. Mutações no gene MAPT resultam em doenças como a Demência frontotemporal (DFT), mas não DA. Alternativamente, os fragmentos Aβ e a isoforma ApoeE4, agem sobre a forma selvagem da proteína Tau, alterandosua função e estrutura e influenciando na fisiopatologia da DA (Imagens do Artigo de HUANG & MUCKE, 2012).

Os tecidos nervosos dos pacientes com DA é caracterizado pelo o excessivo acúmulo do fragmento Aβ, que é insolúvel e possui maior propriedade de agregação (Figura 5). Juntamente com proteínas sinápticas, proteínas inflamatórias, fibrilas neuríticas, células gliais, entre outros componentes formam as placas senis desencadeando toxicidade e morte neuronal (BLACKER &TANZI, 1998; HUANG & MUCKE, 2012).

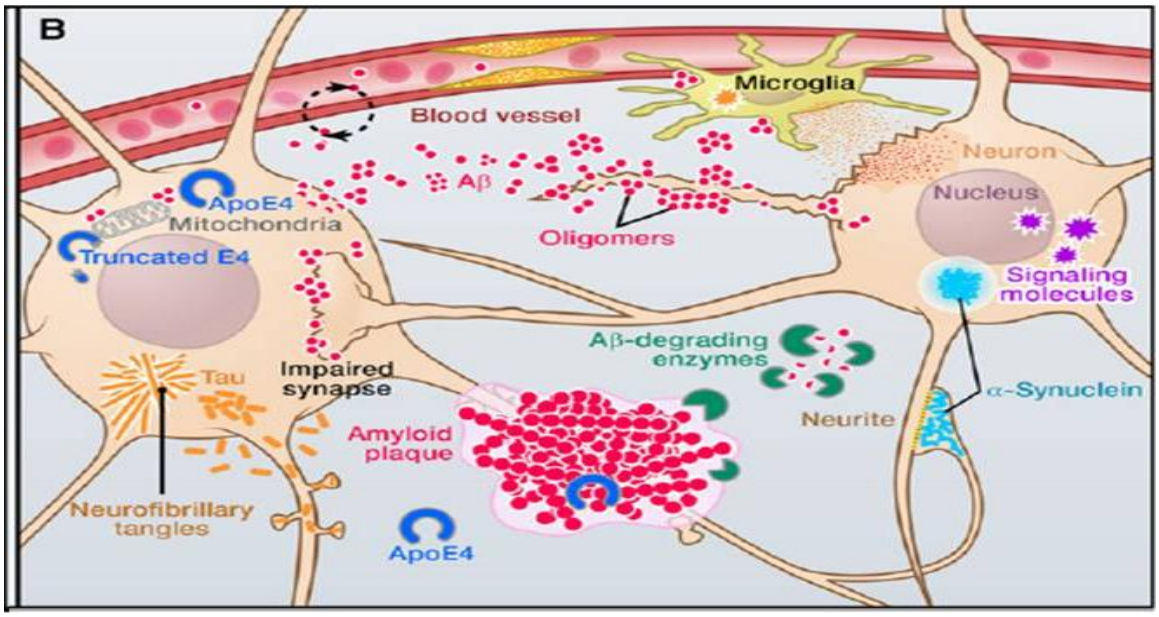



Figura 5. Esquema representativo da agregação e acúmulo do fragmento Aβ. O fragmento causa prejuízo sobre as funções sinápticas e vias de sinalização, a isoforma APOE4 atua impedindo o *clearance* e promove à deposição do fragmento Aβ. A proteína Tau e α-sinucleína podem também ser liberadas para o espaço extracelular e assim espalhar-se a outras células vasculares.(Imagens do artigo de HUANG & MUCKE, 2012).

3.2 Neuroinflamação x DA

A geração e acúmulo dos resíduos ($A\beta_{40-42}$) e dos emaranhados neurofibrilares são os iniciais mecanismos envolvidos com o processo patogênico da DA. Evidências sugerem que o processo inflamatório seria o terceiro mecanismo envolvido, que uma vez iniciado pela degeneração, poderia contribuir para a progressão da doença ( HENEKA & O'BANION, 2007; WEITZ & TOWN, 2012).

Polimorfismos genéticos, fatores epigenéticos (presença de outras doenças), infecções microbianas, injurias traumáticas e uso de drogas, podem ser responsáveis pela transição de uma resposta infamatória benéfica para um processo acelerado de neuroinflamação, além disso, respostas diferenciadas podem estar associadas a diferentes níveis de depósitos das placas *B*-amilóide, de emaranhados neurofibrilares e a presença ou ausência do alelo ε4 APOE (IQBAL, 2005).

Outros estudos demonstram que há classes diferenciadas de microglias, que respondem bem ou mal ao processo neuroinflamatório da DA, sendo dois fatores responsáveis pela qualificação dessa reposta como boa ou má: o estado de dobramento da microglia e a fase da resposta inata se ocorre antes ou depois da inflamação estabelecida (Figura 6) relaciona o estado de dobramento de classes de microglias com sua funcionalidade (WEITZ & TOWN, 2012).

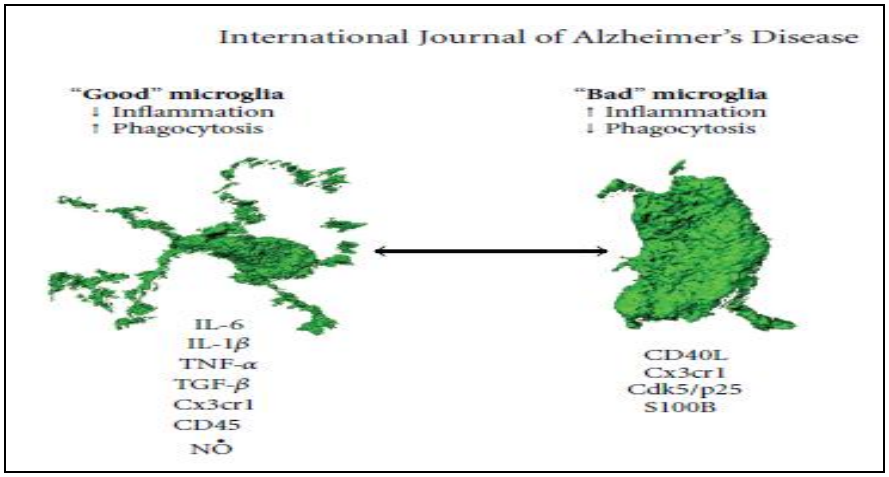



Figura 6. Ilustração 3D do modelo da "good" ramificada e "bad" ovóide microglia. Acredita-se que fenotipicamente a microglia pode oscilar entre essas duas formas. Essas imagens 3D foram geradas através do software Imaris Bitplane 3D cortesia de David Gate. (Imagem do artigo de WEITZ & TOWN, 2012).

3.3 Tipos de estudos usados para a identificação de fatores de riscos genéticos na Doença de Alzheimer.

Estudo de ligação (*Linkage*), de Associação Caso Controle e GWAS - *Genome Wide Association Studies* geralmente são usados na investigação de fatores de risco genéticos para a DA, no entanto algumas considerações são relevantes para cada estratégia dessa.

Esses estudos de ligação geralmente não identificam gene ou mutação, mas sim regiões cromossômicas, todavia esses mapas de ligações têm sido bastante úteis na investigação genética da DA (GUERREIRO *et al*., 2010). Em um importante estudo de ligação Bluter *et al*. investigaram (2.206) afetados e (785) famílias e evidenciaram três regiões candidatas, sendo a região 8q 22-p21.1 portadora do loco que inclui o gene CLU, um dos dez principais genes mais associados a DA, evidenciado em um grande estudo de GWAS (BLUTER *et al*., 2009).

Análises de associação caso controle investigam algumas variantes em um ou dois genes, quando se aumenta o número de variantes e genes, simultaneamente se reduz as chances de achados verdadeiramente positivos. Mesmo assim, muitos desses estudos ainda são constantemente realizados e alternativamente os bancos de dados que reúnem esses achados estão usando critérios mais refinados de meta análise para bem filtrar os achados falsos positivos e falsos negativos (GUERREIRO *et al*., 2010).

No grande acervo de dados do Alzgene alguns desses achados foram confirmados independentemente, demonstrando que essa condição é poligênica e com um padrão de herança genético complexo (BERTRAN *et al*., 2007, 2008).

Estudos de GWAS usados tradicionalmente para identificar fatores de riscos comuns, permitem analisar uma maior quantidade de genes em uma escala de milhares indivíduos casos e controles, através de plataformas como Celera, Afflymetrix (500K) e Illumina.

O sucesso desses estudos depende do tamanho da amostra, da freqüência dos alelos de riscos e do "tamanho do efeito individual". Variantes raras (com freqüência alélica menor que 5% ou mesmo 1%) são mais difíceis de serem identificadas comparadas com variantes comuns ligada ao risco moderado ou leve dentro da população. Para as doenças complexas apenas uma pequena fração dos traços herdáveis podem ser explicados por variações genéticas (STRINGER *et al*., 2011).

Recentes análises de GWAS envolvendo milhares de casos e controles em duas fases de estudos apontam com uma significativa associação três genes além do APOE ligados a DA, são eles: o *CLU* no cromossomo 8 (rs11136000), *PICALM* cromossomo

11(rs3851179), *CR1* cromossomo1q32. Esses três genes fazem parte de uma nova via para progressão da DA. (HAROLD *et al*., 2009; LAMBERT *et al*., 2009).

4. Estudos de Sequenciamentos

O sequenciamento em larga escala do genoma completo de indivíduos afetados tem sido uma abordagem bastante usada para identificar mutações. No entanto, o custo desse experimento ainda é muito elevado e usualmente pouco praticado quando se trata de doenças raras de origem mendeliana, onde alguns fatores são limitantes, tais como; pequeno número de indivíduos afetados ou famílias, heterogeneidade de lócus ou reduzida reprodutividade dos achados (KU *et al*., 2011).

Alternativamente, sequenciamentos de exomas (conjunto de exons do genoma) vem se tornando uma estratégia mais eficiente e de menor custo. Essa metodologia permite a identificação de novas mutações patogênicas, a partir de famílias nucleares, com poucos afetados ou mesmo no estudo de casos isolados de famílias não relacionadas (NG *et al*., 2009; 2010).

Os projetos de resenquenciamento, genotipagem, genomas personalizados, tem gerado uma enorme quantidade de dados genéticos. A bioinformática tem uma importância fundamental na agilização dessas análises, permitindo investigar múltiplos dados simultaneamente, usando algoritmos complexos e auxiliando no melhor entendimento de diversas doenças de base genética, principalmente as de heranças multifatoriais e poligênicas como ocorre com a DA (PEREZ-IRATXETA *et al*., 2002; JOHNSON, 2009).

Existem cerca de 4 milhões de SNPs já identificados no genoma humano "HapMap Project" e com previsão de outras seqüências a serem identificadas. Os estudos de fatores de riscos genéticos apontam os SNPs como as mais freqüentes fontes de variações que predispõem à forma tardia da DA (FREUDENBERG-HUA *et al*., 2003; PSYCHIATRIC GWAS CONSORTIUM COORDINATING COMMITTEE, 2009; www.alzgene.com).

Inserções e Deleções representam o segundo maior grupo de classes de variações genéticas depois dos SNPs, e vários estudos apontam pequenos INDELs como marcadores associados a doenças neuropsiquiátricas, tais como: Alzheimer, Esquizofrenia, Autismo e Retardo Mental (LEMOS *et al*., 2012a; KU *et al*., 2010; LEVY *et al*., 2007).

Algumas abordagens têm sido propostas para se detectar SNPs e INDELs, no entanto a Polymerase Chain Reaction (PCR) seguido do sequenciamento automático direto do DNA, são métodos bastante efetivos para identificação dessas variações (MULLANEY *et al*., 2010).

As plataformas NGS são capazes de gerar informações de até milhões de pares de bases em única corrida. Plataformas como 454 FLX/Roche e Illumina realizam paralelamente amplificação e sequenciamento do DNA. No entanto, algumas divergências entre número de INDELs por genoma individual são relatadas, essa discordância pode ser explicada pelas diferentes abordagens usada por cada técnica (LEMOS *et al*., 2012a).

## 4.1 Bioinformática como ferramenta na identificação de fatores de risco genético

A bioinformática permite integrar informações de diferentes bancos genéticos, e assim desenvolver novas metodologias a partir desses dados, dando direcionamento e especificidade a uma determinada investigação. Todo esse cenário impulsionou o desenvolvimento de um *pipeline*, pelo nosso grupo, com o objetivo de investigar variações genéticas, associadas às doenças neuropsiquiátricas (LEMOS *et al*., 2009; SOUZA *et al*., 2010).

Nessa abordagem os estudos de *Microarrays* de expressão seriam as fontes para a seleção de genes candidatos, e as potenciais novas variações genéticas, seriam identificadas a partir de alinhamentos com o banco de dados público de Expressed Sequence Tags (ESTs). Todo os alinhamentos foram feitos pelo software CLCbio Workbench Combined® (LEMOS *et al*., 2009; SOUZA *et al*., 2010).

Foram identificadas classes de variações genéticas, principalmente deleções que variaram de 1 a 10pb, nas regiões exônicas de grupo de genes associados a DA, a Esporádica Doença de Creutzfeldt-Jakob (CJD) e aos Transtornos do Humor (LEMOS *et al*., 2009; SOUZA *et al*., 2010).

### 4.1.1 *Screening* de deleções de 1-10pb em amostras de DA da população Brasileira

O projeto inicial da validação experimental desse *pipeline* foi realizado, através de sequenciamentos, de amostras de pacientes Brasileiros com DA, de algumas variações preditas *in silico*.

A metodologia usada foi segundo LEMOS *et al*. (2011), os primers foram desenhados através do programa Primer3 as amostras de DNA foram amplificadas por PCR e sequenciadas pela plataforma Mega Bace 1000 (Sunnyvale, CA).

Os resultados revelaram alguns falsos positivos, sendo meros artefatos da técnica, que logo foram descartados após o sequenciamento da fita complementar. Foram avaliados três genes ligados à via inflamatória da DA, sendo eles: COX2 (exon 1 e 7), CD83 (exon 5), IL1A (exon 5), onde duas variações do tipo SNPs: "rs2745557"_COX2 (exon 1) e "rs17561"_IL1A (exon5), foram encontradas. Os preliminares experimentos não detectaram nenhuma das novas mutações candidatas (Figura 7).

O desafio atual é melhorar o sequenciamento, através de diferentes abordagens, a fim de detectar esse tipo de variação.

Em técnicas forenses, o uso de plataformas multiplex e eletroforese *dye-labeling*, são abordagens utilizadas para detecção de pequenos INDELs (PEREIRA *et al*. 2009a; PEREIRA *et al*. 2009b).

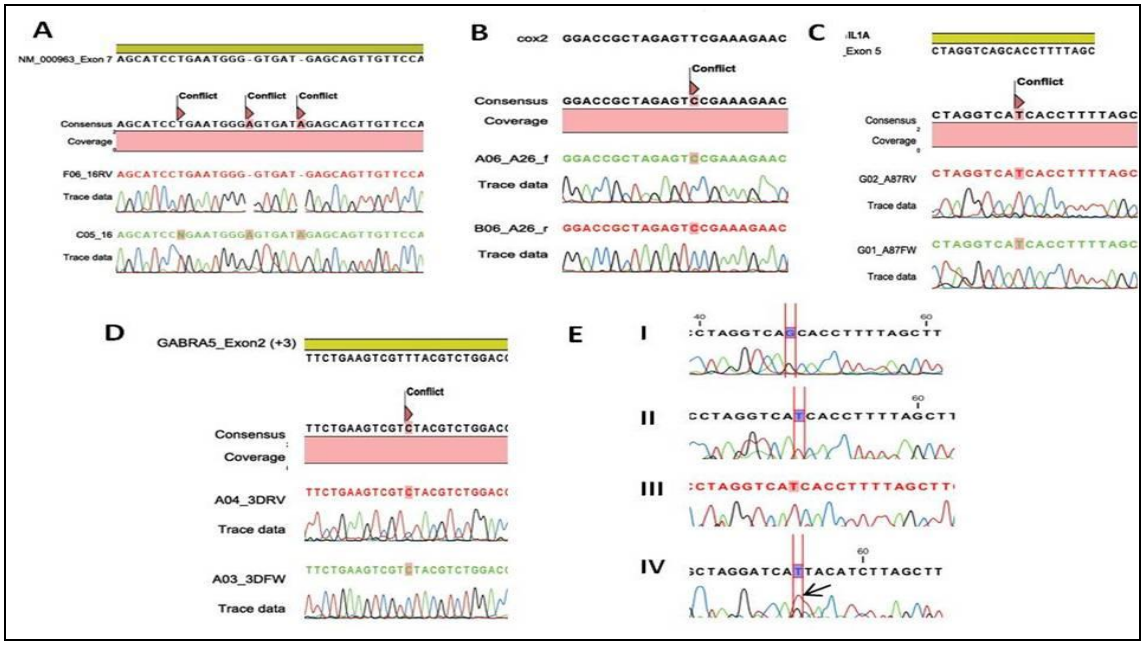



Figura 7. Sequenciamentos realizados pela plataforma Mega Bace 1000. Os conflitos representam variações. **A**. Mostram alguns falsos positivos identificados. **B**, **C**, **D**. São SNPs validados e depositados no dbSNP/NCBI para os genes COX2, IL1A e GABRA5. **E**. Demonstra um experimento de DNA *pooling* com três amostras, foram usados 0,33µl de DNA de cada amostra. A seta indica o pico do pool.

## 5.0 Referências Bibliográficas

Associação Brasileira de Alzheimer <//http:www.abraz.br>. Acesso: 20 de junho 2012.

ARGYELAN M, CARBON M, NIETHAMMER M et al. Cerebellothalamocortical connectivity regulates penetrance in dystonia. J Neurosci,v.31, p.9740-9747, 2009.

BARROS E SILVA E. Prevalence analysis of basal ganglia calcification using neuroimaging. Dissertação de Mestrado. Federal University of Pernambuco, 2008.

BERTRAM, L.; MCQUEEN, M. B.; MULLIN, K et al. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet, v.39, 17–23, 2007.

BERTRAM, L.; TANZI, R. E. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. Nat Rev Neurosci, v. 9, 768–778, 2008.

BLACKER, D.; TANZI, R. E. The genetics of Alzheimer disease: current status and future prospects. Arch Neurol, v.55, p. 294-296, 1998.

BUTLER, A.W.; NG, M.Y.; HAMSHERE, M.L et al. Meta-analysis of linkage studies for Alzheimer's disease _ a web resource. Neurobiol Aging, v. 30, p. 1037–1047, 2009.

CHAU, D. M.; CRUMP, C. J.; VILLA, J. C.; SCHEINBERG, D. A. Familial Alzheimer disease presenilin-1 mutations alter the active site conformation of γ-secretase.J Biol Chem. 2012 May 18;287(21):17288-96. Epub 2012 Mar 29.

CRUTS, M.; THEUNS, J.; VAN BROECKHOVEN, C. Locus-Specific Mutation Databases for Neurodegenerative Brain Diseases. Human Mutation, v.00, p.1-5, 2012.

DAI, X.; GAO, Y.; XU, Z. et al. Identification of a novel genetic locus on chromosome 8p21.1-q11.23for idiopathic basal ganglia calcification. Am J Med Genet B Neuropsychiatr Genet, v.153, p. 1305-10, 2010.

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

DUBOIS, B.; FELDMAN, H.H.; JACOVA, C et al. Research criteria for the diagnosis of Alzheimer's Disease: Revising the nincds–adrda criteria. Lancet Neurol, v. 6, p. 734–746, 2007.

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

FUJITA, D.; TERADA, S.; ISHIZU, H et al. Immunohistochemical examination on intracranial calcification in neurodegenerative diseases. Acta Neuropathol (Berl), v.3, p.259-64, 2003.

GESCHWIND, D. H.; LOGINOV, M.; STERN, JM. Identification of a locus on chromosome 14q for idiopathicbasal ganglia calcification (Fahr disease). Am J Hum Genet, v.65, p. 764-72, 1999.

GIACHELLI, C.M. The Emerging Role of Phosphate in Vascular Calcification. Kidney Int, v, 75, p.890-897, 2009.

GOLDMAN, J. S. New Approaches to Genetic Counseling and Testing for Alzheimer's Disease and Frontotemporal Degeneration. Curr Neurol Neurosci Rep *DOI 10.1007/s11910-012-0296-1*, 2012.

GUERREIRO, R. J.; BAQUERO, M.; BLESA, R et al. Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP. Neurobiol Aging, v. 31, p.725–731, 2010.

HAROLD, D.; ABRAHAM, R.; HOLLINGWORTH, P et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet, v. 41, 1088–1093, 2009.

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

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

HSU, C. S.; SEARS, R.; LEMOS, R. R et al. Mutations in *SLC20A2* are a Major Cause of Familial Idiopathic Basal Ganglia Calcification. Neurogenetics, 2012 (no prelo).

HUANG, Y.; MUCKE, L. Alzheimer mechanisms and therapeutic strategies. , v.148, p.1204-1222, 2012.

HUANG, Y. Molecular and cellular mechanisms of apolipoprotein E4 neurotoxicity and potential therapeutic strategies. Curr. Opin. Drug Discov. Devel, v.9, p.627–64, 2006.

IQBAL, K.; FLORY, M.; KHATOON, S et al. Subgroups of Alzheimer's disease based on cerebrospinal fluid molecular markers. Ann. Neurol, v.58, p. 748–757, 2005.

JOHNSON, A. D. SNP bioinformatics: a comprehensive review of resources. Circ Cardiovasc Genet, v. 5, p. 530-536, 2009. Doi:10.1161/CIRCGENETICS.109.872010.

KAVANAUGH, M. P.; KABAT, D. Identification and characterization of a widely expressed phosphate transporter/retrovirus receptor family. Kidney int, v. 49, p.959 –63, 1996.

KEMPTOM, M. J.; HALDANE, M.; JOGIA, J et al. Dissociable brain structural changes associated with predisposition, reliance, and disease expression in bipolar disorder. J. Neurosci, v.29, p.10863-10868, 2009.

KENDRICK, J., KESTENBAUM, B. & CHONCHOL, M. Phosphate and cardiovascular disease. Adv. Chronic Kidney Dis, v. 18, p.113–119, 2011.

KIM, J.; BASAK, J. M.; HOLTZMAN, D. M. The role of apolipoprotein E in Alzheimer's disease. Neuron, v. 63, p. 287–303, 2009.

KU, C. S.; NAIDOO, N.; PAWITAN, Y. Revisiting Mendelian disorders through exome sequencing. Hum Genet, v. 129, p.351–70, 2011.

LAMBERT, J. C.; HEATH, S.; EVEN, G et al. Genome-Wide Association Study Identifies Variants at CLU and CR1 Associated With Alzheimer's Disease. Nat Genet, v. 10, p. 1094–1099, 2009.

LAGRUE, E.; ABE, H.; LAVANYA, M et al. Regional characterization of energy metabolism in the brain of normal and MPTP-intoxicated mice using new markers of glucose and phosphate transport. Journal of Biomedical Science, v. 17, p.91, 2010.

LEVY, S.; SUHON, G.; NG, P. C et al. The diploid genome sequence of an individual human. Plos Biol, v.10: e254, 2007.

LEMOS, R. R.; CASTELLETTI, C. H.; LIMA FILHO, J. L et al. In silico identification of new genetic variations as potential risk factors for Alzheimer's disease in a microarray-oriented simulation. J Mol Neurosci, v. 39, p. 242-247, 2009.

LEMOS, R. R.; OLIVEIRA, D. F.; ZATZ, M.; OLIVEIRA, J. R. M. Population and Computational Analysis of the MGEA6 P521A Variation as a Risk Factor for Familial Idiopathic Basal Ganglia Calcification (Fahr's Disease). Journal of Molecular Neuroscience, v. 43, p. 333-336, 2011.

LEMOS, R. R.; OLIVEIRA, M.F.; OLIVEIRA, J. R. M. Reporting a new mutation at the SLC20A2 gene in a Brazilian family with idiopathic basal ganglia calcification. European Journal of Neurology, 2012b (no prelo).

LEMOS, R. R.; SOUZA, M. B.; OLIVEIRA, J. R. Exploring the implications of INDELs in neuropsychiatric genetics: challenges and perspective. J Mol Neurosci, v. 47, p. 419-424, 2012a.

LI, G., BIEN-LY, N., ANDREWS-ZWILLING, Y et al. GABAergic interneuron dysfunction impairs hippocampal neurogenesis in adult apolipoprotein E4 knockin mice. Cell Stem Cell, v.5, p. 634–645, 2009.

MANYAM, B. V. What is and what is not 'Fahr's disease'. Parkinsonism Relat Disord, v. 11, p.73-80, 2005.

MORRIS, J.C., ROE, C.M., GRANT, E.A., et al. Pittsburgh compound B imaging and prediction of progression from cognitive normality to symptomatic Alzheimer disease. Arch. Neurol, v. 66, p. 1469–1475, 2009.

MULLANEY, J. M.; MILLS, R. E.; PITTARD, W.S.; DEVINE, S. E. Small insertions and deletions (INDELs). Hum Mol Genet, v.19, p.131-136, 2010.

NG, S. B.; TURNER, E. H.; ROBERTSON, P. D et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature, v. 461, p.272–76, 2009.

NG, S. B.; BIGHAM, A. W.; BUCKINGHAM, K. J et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet ,v 42, p.790–93, 2010.

OLIVEIRA, J. R. M. Managing Idiopathic Basal Ganglia Calcifications (“Fahr’s Disease”). 1ºed. Nova Science Publishers, New York, 2011.

OLIVEIRA, J. R.M.; HOPFER, S.; PAPP, J. et al. A genome wide scan for familial idiopathic basal ganglia calcification (Fahr’s disease) identifies new candidate regions and confirms genetic heterogeneity. American Journal Human Genetic, v. 5, p.570, 2003.

OLIVEIRA, J.R.M.; LEMOS R.R. Linkage studies in familial idiopathic basal ganglia calcification: Separating the wheat from the chaff. American Journal of Medical Genetics. Part B, v. 150B, p. 594-595, 2009c.

OLIVEIRA, J. R. M.; LEMOS, R. R.; OLIVEIRA, M. F . Updating genetics studies in familial idiopathic basal ganglia calcification. Southern Medical Journal (Birmingham, Ala. Print), v. 102, p. 989-989, 2009b.

OLIVEIRA, J. R. M.; LIMA FILHO, J. L.; ZATZ, M. Identical twins with Idiopathic Basal Ganglia Calcification ( Fahr Disease ) presenting with a remarkably similar

pattern of neuroimaging findings. Parkinsonism & Related Disorders, v. 15, p. 396-97, 2009d.

OLIVEIRA, J. R. M.; OLIVEIRA, M. F.; LEMOS, R. R.; OLIVEIRA, M. L . The current status of "Fahr's disease" nosology. Practical Neurology, v. 9, p. 302-302, 2009a.

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

OLIVEIRA, JRM.; SOBRIDO, MJ.; SPITERI, E et al. Analysis of candidate genes at the IBGC1 locus associated with idiopathic basal ganglia calcification ("Fahr's Disease"). J Mol Neurosci, v. 33, p.151-54, 2007.

OLIVEIRA, J. R. M.; SPITERI, E.; SOBRIDO, M. J et al. Genetic heterogeneity in idiopathic basal ganglia calcification (Fahr disease). Neurology, v, 63, p. 2165-67, 2004.

PASCHALI, A.; LAKIOTIS, V.; MESSINIS, L et al. Dopamine transporter SPECT/CT and perfusion brain SPECT imaging in idiopathic basal ganglia calcinosis. Clin. Nucl. Med, v. 34, p. 421-423, 2009.

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

PEREIRA, R.; PHILLIP, C.;ALVES, C et al. Insertion/deletion polymorphisms: a multiplex assay and forensic applications. Forensic Sci Int Genet Suppl, v. 2, p.513–515, 2009a.

PEREIRA, R.; PHILLIPS, C.; ALVES, C et al. A new multiplex for human identification using insertion/ deletion polymorphisms. Electrophoresis, v. 30, p.3682–3690, 2009b.

PSYCHIATRIC GWAS CONSORTIUM COORDINATING COMMITTEE. Genomewide Association Studies: History, Rationale, and Prospects for Psychiatric Disorders Psychiatric GWAS Consortium Coordinating Committee. American Journal Psychiatry, v.166, p.540-556, 2009. doi: 10.1176/appi.ajp.2008.08091354

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

SAIKI, M.; SAIKI, S.; MATSUNARI, I et al. Neurological deficits are associated with increased brain calcinosis, hypoperfusion, and hypometabolism in idiopathic basal ganglia calcification. Mov. Disord, v.7, p.1027-1030, 2007.

SÁNCHEZ-PÉREZ, A. M.; CLARAMONTE-CLAUSELL, B.; SÁNCHEZ-ANDRÉS J. V.; HERRERO, M. T. Parkinson's disease and autophagy. Parkinsons Dis, v. doi: 10.1155/2012/429524. Epub 2012 Oct 17.

SOUZA, M. B.; LEMOS, R. R.; CUNHA, J. E et al. Searching for new genetic risk factors for neuropsychiatric disorders in expression databases. J Mol Neurosci, v. 41, p.193-197, 2010.

SMALL, G. W.; ERCOLI, L. M.; SILVERMAN, D. H, et al. Cerebral metabolic and cognitive decline in persons at genetic risk for Alzheimer's disease. Proc Natl Acad Sci U S A, v. 97, p. 6037-6042, 2000.

SOBRIDO, M. J.; HOPFER, S.; GESCHWIND, D. H. Familial Idiopathic Basal Ganglia Calcification. In Pagon et al, editors. GeneReviews: Genetic Disease Online Reviews Seattle, updated 2007. www.geneclinics.org.

STRINGER, S.; WRAY, N. R.; KAHN, R. S.; DERKS, E. M. Underestimated Effect Sizes in GWAS: Fundamental Limitations of Single SNP Analysis for Dichotomous Phenotypes. PLoS ONE, v.11, e27964. doi:10.1371/journal.pone.0027964, 2011.

VOLPATO, C. B.; DE GRANDI, A.; BUFFONE, E. Exclusion of linkage to chromosome 14q in a large South Tyrolean family with Idiopathic Basal Ganglia Calcification (IBGC). Am J Med Genet B Neuropsychiatr Genet. 147, p. 1319-1922, 2008.

VOLPATO, CB.; DE GRANDI, A.; BUFFONE, E et al. 2q37 as a susceptibility locus for idiopathic basalganglia calcification (IBGC) in a large South Tyrolean family. J Mol Neurosci,v.39, p.346-53, 2009.

VILLA-BELLOSTA, R.; RAVERA, S.; SORRIBAS, V et al. The Na-Pi cotransporter PiT-2 (SLC20A2) is expressed in the apical membrane of rat renal proximal tubules and regulated by dietary Pi. *Am. J. Physiol.* Renal Physiol, v. 296, p. 691–99, 2009.

WANG, C.; LI, Y.; SHI, L et al. Mutations in *SLC20A2* link familial idiopathic basal ganglia calcification with phosphate homeostasis. Nature Genetics, v. 44 p.254-256, 2012.

WEITZ , T.M.; TOWN,T. Microglia in Alzheimer's Disease: It's All About Context. Int J Alzheimers Dis. 2012;2012:314185. Epub 2012 Jun 20.

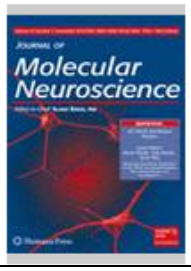
WINKLHOFER, K. F.; TATZELT, J.; HAASS, C. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. EMBO J, v. 27, p. 336-349, 2008.

World Health Organization< http://www.who.int> Acesso em 04 de Novembro 2012

YANKNER, B.A.; LU, T.; LOERCH, P. The aging brain. Annu Rev Pathol.v. 3, p.41–66, 2008.

# Capítulo 2

JOURNAL OF MOLECULAR NEUROSCIENCE
Volume 47, Number 3 (2012), 419-424, DOI: 10.1007/s12031-012-9714-8




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

R. R. Lemos, M. B. R. Souza and J. R. M. Oliveira

J Mol Neurosci
DOI 10.1007/s12031-012-9714-8

# 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


**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 microdeletions 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

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

Published online: 16 February 2012

## 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

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 (Syvänen 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 overage 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, 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

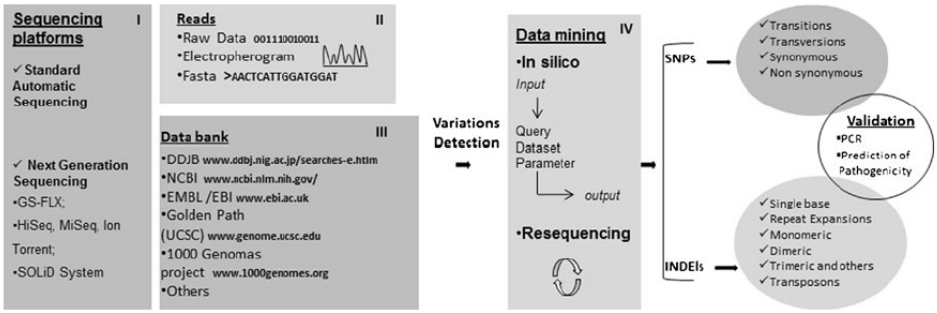



**Fig. 1** A pipeline suggested to merge different data resources in order to combine expression studies and genetic variations and predicted to detect potential new polymorphisms, including INDELs

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

Arnbruster 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, Kretzschmar 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 M, Du J, Gerstein M (2010) Personal genome sequencing: current approaches and challenges. Genes & Devepment 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
Togsverd 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 3

JOURNAL OF MOLECULAR NEUROSCIENCE
Volume 43, Number 3 (2011), 333-336, DOI: 10.1007/s12031-010-9445-7


## Population and Computational Analysis of the MGEA6 P521A Variation as a Risk Factor for Familial Idiopathic Basal Ganglia Calcification (Fahr's Disease)

Roberta R. Lemos, Danyllo F. Oliveira, Mayana Zatz and João R. M. Oliveira

J Mol Neurosci (2011) 43:333–336
DOI 10.1007/s12031-010-9445-7

# Population and Computational Analysis of the MGEA6 P521A Variation as a Risk Factor for Familial Idiopathic Basal Ganglia Calcification (Fahr's Disease)

**Roberta R. Lemos · Danyllo F. Oliveira · Mayana Zatz · João R. M. Oliveira**

Received: 17 August 2010 / Accepted: 26 August 2010 / Published online: 14 September 2010


**Abstract** Familial idiopathic basal ganglia calcification, also known as "Fahr's disease" (FD), is a neuropsychiatric disorder with autosomal dominant pattern of inheritance and characterized by symmetric basal ganglia calcifications and, occasionally, other brain regions. Currently, there are three loci linked to this devastating disease. The first one (IBGC1) is located in 14q11.2-21.3 and the other two have been identified in 2q37 (IBGC2) and 8p21.1-q11.13 (IBGC3). Further studies identified a heterozygous variation (rs36060072) which consists in the change of the cytosine to guanine located at MGEA6/CTAGE5 gene, present in all of the affected large American family linked to IBGC1. This missense substitution, which induces changes of a proline to alanine at the 521 position (P521A), in a proline-rich and highly conserved protein domain was considered a rare variation, with a minor allele frequency (MAF) of 0.0058 at the US population. Considering that the population frequency of a given variation is an indirect indicative of potential pathogenicity, we screened 200 chromosomes in a random control set of Brazilian samples and in two nuclear families, comparing with our previous analysis in a US population. In addition, we accomplished analyses through bioinformatics programs to predict the pathogenicity of such variation. Our genetic screen found no P521A carriers. Polling these data together with the previous study in the USA, we have now a MAF of 0.0036, showing that this mutation is very rare. On the other hand, the bioinformatics analysis provided conflicting findings. There are currently various candidate genes and loci that could be involved with the underlying molecular basis of FD etiology, and other groups suggested the possible role played by genes in 2q37, related to calcium metabolism, and at chromosome 8 (NRG1 and SNTG1). Additional mutagenesis and in vivo studies are necessary to confirm the pathogenicity for variation in the P521A MGEA6.

**Keywords** MGEA6 · P521A variation · Polymorphism · Familial idiopathic basal ganglia calcification · Fahr's disease

Roberta R. Lemos and Danyllo F. Oliveira contributed equally to this work.

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

M. Zatz
Human Genome Study Center, University of São Paulo,
São Paulo, São Paulo, Brazil

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

## Introduction

Familial idiopathic basal ganglia calcification (IBGC), also known as "Fahr's disease" (FD), is a complex condition which exhibits various neuropsychiatric symptoms such as psychosis, motor impairment, and mood disorders. However, the most striking signal of this condition is bilateral brain calcinosis in basal ganglia, and, occasionally other brain regions, with normal biochemical and endocrinological profiles (Manyam 2005).

The first Fahr's disease locus (IBGC1) was described in a large multigenerational American family in a 13.3-cM region (14q11.2-21.3) with an autosomal dominant pattern of inheritance. Another small kindred from Spain was also

reported as being possibly linked to this locus, narrowing the candidate region to 10.9 cM. (Geschwind et al. 1999; Oliveira et al. 2004).

However, this locus was excluded in other families from China, Canada, and Germany, suggesting genetic heterogeneity (Oliveira et al. 2004). So far, other two loci have been identified in families from Italy (Volpato et al. 2009) and from China (Dai et al. 2010). These are in 2q37 and 8p21.1-q11.23, respectively.

In a systematic search for a candidate mutation at the IBGC1 locus, we identified a heterozygous non-synonymous single nucleotide polymorphism (C>G) at the MGEA6/cTAGE5 gene (rs36060072) in every affected member, but not at the controls, of the family firstly associated with this locus. This missense substitution, which is responsible for the change of the proline to alanine at the 521 position (P521A) in a proline-rich and highly conserved domain, was considered a rare variation, with a minor allele frequency of 0.0058 in the US population (Oliveira et al. 2007).

Curiously, this variation is located at the exon which is commonly spliced at the MGEA6 gene, generating the isoform MGEA 11, also expressed in the brain (Usener et al. 2003). MGEA6 is a coil-coiled protein expressed in several tissues including the brain, highly expressed in meningioma, a commonly benign intracranial tumor habitually presenting calcification visible at neuroimaging studies (Comtesse et al. 2001, 2002; Usener et al. 2003).

Considering that the population frequency of a given variation is an indirect indicative of potential pathogenicity (Freudenberg-Hua et al. 2003), we present now the sequencing of 100 control samples from Brazil to ascertain a local minor allele frequency (MAF) for this variation at the MGEA6, comparing the results with our previous analysis in the USA. We also screened two nuclear families for this polymorphism. One of them has already been reported in literature and presents a 53-year-old male proband with IBGC and neuropsychiatric symptoms (Oliveira et al. 2009). This proband is the father of two 23-year-old asymptomatic identical twins with extensive brain calcinosis in basal ganglia, white matter, and cerebellum. The second family is inedit and also presents a similar pattern of calcification in cerebellum and basal ganglia over three generations, one affected (87 years old) and two unaffected (33 and 58 years old), all females. Altogether, these patients did not present any known metabolic disturbances on clinical examination (Fig. 1).

## Methods and Sample Collection

The DNA was extracted by salting out procedure of blood samples, followed by proteinase K digestion and precipitation with ethanol. The polymorphic region was amplified by polymerase chain reaction (PCR) and the cycling program was 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 51.5°C for 30 s, 72°C for 45 s, followed by a final extension of 72°C for 10 min. The primers, manually designed, used in this procedure were 5′tttcagatgaaggccaacccaatggtgagg3′ (for) and 5′accatatggggaatgctctggaataattaa3′ (rev). PCR was performed in a 25-μl reaction volume containing 100 ng of DNA in a reaction mixture containing 2 μM of $MgCl_2$, 2 μL of dNTPs, 0.75 U *Taq* DNA polymerase (Invitrogen), 1× PCR buffer (Invitrogen) 0.2 μM.

The PCR products were purified by enzymatic reaction of exoquinase/shrimp alkaline phosphatase (GE). Finally, the products were sequenced in a Mega Bace 1000 (Sunnyvale, CA) and ABI Prism 3100 genetic analyzer (Lincoln, CA).

Besides the sequencing analysis, we accomplished an in silico study through bioinformatics programs to perform multiple amino acid sequence alignments and to predict the pathogenicity of a given polymorphism based on the protein primary sequence or tertiary structure, or both. Because the three-dimensional structure of the MGEA6

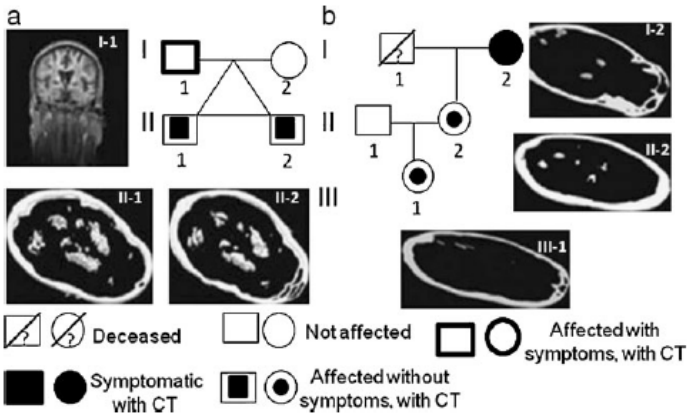



Fig. 1 Analysis of pedigrees in two nuclear families. a The first family has already been reported in literature and has two identical twins with bilateral brain calcification, but without any psychiatric symptoms at all. b The second family has a three-generational transmission of brain calcification and is inedit in literature

**Table 1** Prediction of the pathogenicity and thermodynamic changes caused by P521A MGEA6 polymorphism

| Programs | Score | Prediction | Electronic address |
|---|---|---|---|
| SAP | 0.649 | Pathogenic | http://yanbioinformatics.cs.usu.edu:8080/SAPsubmit |
| PolyPhen | 2020 | Probably damaging | http://genetics.bwh.harvard.edu/pph/ |
| SIFT | −0.88 | Dubious | http://blocks.fhcrc.org/sift/SIFT.html |
| I-Mutant | | Decreased molecular stability | http://gpcr.biocomp.unibo.it/~emidio/I-Mutant/I-Mutant.htm |
| Pmut | 0.3756 | Not pathogenic | http://mmb2.pcb.ub.es:8080/PMut/ |

genic product is not yet available, we focused only in the examination through the sequence.

Initially, the software CLCBio Workbench Combined®, version 3.6.2, was used to overlap and compare the candidate region where the P521A was found, between the human MGEA6 and its orthologous genes from *Pan troglodytes*, *Canis lupus familiaris*, *Sus scrofa*, *Mus musculus*, *Gallus gallus*, and *Danio rerio*.

The bioinformatics analysis was performed with five programs: PMUT, PolyPhen, SAP, SIFT, and I Mutant. This last one does not predict if a determined mutation is pathogenic, but analyzes the changes in the free energy levels caused by the mutation and if this event might disrupt the structural stability of this macromolecule.

This study was approved by the ethics committee from the Federal University of Pernambuco (CAEE-0296.0172.000-08) and the subjects signed informed consent.

## Results and Discussion

We found no P521A carriers during the sequencing screening of the control samples (Table 1). Before our studies, the MAF of this polymorphism was 0.0058 (Oliveira et al. 2007).

There is an important ethnical heterogeneity when comparing random populations between Brazil and the USA; however, polling these data together with the previous study in the USA, we have now a MAF of 0.0036. In previous studies, it was attested that possible damaging mutations have a MAF<0.05 in a European population (Freudenberg-Hua et al. 2003). Thus, this result corroborates to point the P521A MGEA6 as a potentially pathogenic mutation. None of the patients from the two FD families presented the polymorphism, reinforcing the genetic heterogeneity to this disease (Oliveira et al. 2004).

The orthologous gene comparison shows that this variation is located in a highly conserved region, suggesting that a non-synonymous mutation at this spot is potentially deleterious (Fig. 2).

The bioinformatics analysis provided conflicting findings (Table 1). SAP, PolyPhen, and I Mutant suggest that this variation is predicted to cause probable molecular deleterious effects. I Mutant is the only program that does not predict the pathogenicity, but rather analyzes the mutation through a thermodynamic perspective, pointing that this polymorphism decreases protein stability. In addition, PolyPhen classified this mutation as "probably damaging."

The SIFT program generated conflicting findings. The use of UniProt-SWISS PROT 56.6, and NCBI databases, differently used during the comparative predictions, pointed this polymorphism as pathogenic. However, the same analysis using the UniProt-TrEMBL 39.6 database considered this substitution as neutral. The PMUT considered this same mutation as neutral.

The growing number of annotated SNP in databases raises the need of in silico tools with a high predictive power because it is unfeasible to verify individually the ramping number of variations found during the current sequencing and resequencing projects worldwide. All of these programs have been reported in the literature as highly precise tools to screen candidate SNPs; however, other studies have pointed potential failures (Ferrer-Costa et al. 2005; Ng and Henikoff 2006; Hu and Yan 2008; Di et al. 2009; Dorfman et al. 2010).

The lack of consensus in the analysis might also be due to the limited skill of these programs in the prediction. In addition, the discrepancies might be due to the poor source of data available for some genes. In the MGEA6 case, there is no structural model to the proteic product, and it could be a restraint to these online programs.


Fig. 2 Multiple alignment of the MGEA6 proline-rich region where the polymorphism is located (*pointed by arrow*) shows that this region is highly evolutionarily conserved

Proline-rich domains are reported as a widespread motif in genomes of various organisms and play a critical role in the construction of preferential three-dimensional structures, recognizing and interacting with domains such as WW and Src homology 3, which are involved in a wide variety of cellular activities such as growth, cell cycle, transcription, synaptic signalization, and cell motility (Zarrinpar et al. 2003).

In addition, there are other suggestive candidate genes that could be the underlying molecular basis of FD etiology. Volpato et al. (2009) point out several genes related to calcium metabolism located in 2q37, such as INPP5 and EFHD1. Curiously, two families with Fried syndrome (X-linked mental retardation with hydrocephalus and calcifications in basal ganglia) have been reported in France and Scotland and bearing AP1S2 mutations (Saillour et al. 2007).

More recently, Dai et al. (2010) reported a Chinese family linked to chromosome 8, but an initial screening at the NRG1 and SNTG1 genes found no mutation.

Previous preliminary linkage studies pointed to other candidate regions at chromosomes 7, 8, and 9, but additional fine mapping analyses are pending (Oliveira et al. 2003; Wszolek et al. 2006)

Therefore, continuous familial and population screening, in vivo studies, and mutagenesis analysis are necessary to definitely comprehend the pathogenesis of Fahr's disease

**Acknowledgments** We are deeply indebted to Dr. Cíntia Rocha, Dr. José Luiz de Lima Filho (LIKA), and Dr Constância Ayres (Aggeu Magalhães Research Center) for the support with sequencing experiments. We also thank Dr. Maria Rita Passos Bueno and Lilian Calabró dos Santos (University of São Paulo) for providing the control DNA samples. This study was supported by the following Brazilian funding agencies and academic bureaus: CNPq (Universal 2008-480149/2008-9), PROPESQ-UFPE, CAPES, and FACEPE (APQ-0997-4.01/08) and PIBIC-UFPE. J.R. M.O. holds a research fellowship from the Brazilian National Research Council (CNPq, Brasília, Brazil) and from the John Simon Guggenheim Foundation (New York, USA).

## References

Comtesse N, Reus K, Meese E (2001) The MGEA6 multigene family has an active locus on14q and at least nine pseudogenes no different chromosomes. Genomics 75:1–6

Comtesse N, Niedermayer I, Glass B et al (2002) MGEA6 is tumor-specific overexpressed and frequently recognized by patient-serum antibodies. Oncogene 21:239–247

Dai X, Gao Y, Xu Z et al. (2010) Identification of a novel genetic locus on chromosome 8p21.1-q11.23 for idiopathic basal ganglia calcification. Am J Med Genet B Neuropsychiatr Genet. doi:10.1002/ajmg.b.31102

Di YM, Chan E, Wei MQ, Liu JP, Zhou SF (2009) Prediction of deleterious non-synonymous single-nucleotide polymorphisms of human uridine diphosphate glucuronosyltransferase genes. AAPS J 11:469–480

Dorfman R, Nalpathamkalam T, Taylor C et al (2010) Do common in silico tools predict the clinical consequences of amino-acid substitutions in the CFTR gene? Clin Genet 77:464–473

Ferrer-Costa C, Gelpí CL, Zamakola L, Parraga I, Cruz X, Orozoco X (2005) PMUT: a web-based tool for the annotation of pathological mutation on proteins. Struct Bioinform 21:3176–3178

Freudenberg-Hua Y, Freudenberg J, Kluck N, Cichon S, Propping P, Nothen MM (2003) Single nucleotide variation analysis in 65 candidate genes for CNS disorders in a representative sample of the European population. Genome Res 13:2271–2276

Geschwind DH, Loginov M, Stern JM (1999) Identification of a locus on chromosome 14q for idiopathic basal ganglia calcification (Fahr disease). Am J Hum Genet 65:764–772

Hu J, Yan C (2008) Identification of deleterious non-synonymous single nucleotide polymorphisms using sequence-derived information. BMC Bioinform 9:1–8

Manyam BV (2005) What is and what is not 'Fahr's disease'. Parkinsonism Relat Disord 11:73–80

Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 7:61–80

Oliveira JRM, Hopfer S, Papp J et al (2003) A genome wide scan for familial idiopathic basal ganglia calcification (Fahr's disease) identifies new candidate regions and confirms genetic heterogeneity. Am J Hum Genet 5:570

Oliveira JR, Spiteri E, Sobrido MJ et al (2004) Genetic heterogeneity in familial idiopathic basal ganglia calcification ("Fahr's disease"). Neurology 63:2165–2167

Oliveira JR, Sobrido MJ, Spiteri E et al (2007) Analysis of candidate genes at the IBGC1 locus associated with idiopathic basal ganglia calcification ("Fahr's disease"). J Mol Neurosci 33:151–154

Oliveira JR, Lima Filho JL, Zatz M (2009) Identical twins with idiopathic basal ganglia calcification ("Fahr's disease") presenting with a remarkably similar pattern of neuroimaging findings. Parkinsonism Relat Disord 5:396–397

Saillour Y, Zanni G, Des Portes V et al (2007) Mutations in the AP1S2 gene encoding the sigma 2 subunit of the adaptator protein 1 complex are associated with syndromic X-linked mental retardation with hydrocephalus and calcifications in basal ganglia. J Med Genet 11:739–744

Usener D, Schadendorf D, Kock J, Dubel S, Eichmuller S (2003) cTAGE: a cutaneous T cell lymphoma associated antigen family with tumor specific splicing. J Investig Dermatol 121:198–206

Volpato CB, De Grandi A, Buffone E et al (2009) 2q37 as a susceptibility locus for idiopathic basal ganglia calcification (IBGC) in a large South Tyrolean family. J Mol Neurosci 39:346–353

Wszolek ZK, Baba Y, Mackenzie IR et al (2006) Autosomal dominant dystonia-plus with cerebral calcifications. Neurology 22:620–625

Zarrinpar A, Bhattacharyya RP, Lim WA (2003) The structure of proline recognition domains. Sci STKE 22:1–10

# Capítulo 4

# Artigo aceito para publicação

## Mutations in *SLC20A2* are a Major Cause of Familial Idiopathic Basal Ganglia Calcification

Sandy Chan Hsu, MS[1], Renee Sears, BA[1], Roberta R Lemos[5], Beatriz Quintáns[2,3], PhD, Alden Huang[1], Elizabeth Spiteri, PhD[1¶], Lisette Nevarez[1§], Zatz M[7], Vladimir Kostić[11], Milena Janković[11], Valerija Dobricić[12], Ivana Novaković[11], Peter Schofield, Kerrie Pierce, Jan Fullerton, Phillip Mitchell, Henry Brodaty, Jill Goldman, Melissa K Maisenbacher, Jennifer Mueller, Matthew Bower, MS, Martin Paucar Arce, MD, Witoon Mitarnun[13], Agnes Linglart, Sheila Simpson, Zofia Miedzybrodzka, PhD, Joanna Jen, MD PhD, Henry Paulson, MD PhD, Brent Fogel, MD PhD, Michele Yang[3], Serge Jaumà[8], Oga Carmona[9], Daniel García-Estevez[10], Catherine Mamah, MD[1§§], Joao RM de Oliveira[5,6], María-Jesús Sobrido, MD PhD[2,4], Daniel H. Geschwind, MD PhD[1,3*], Giovanni Coppola, MD[1,3*]

[1]Program in Neurogenetics, Department of Neurology and [3]Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles CA
[2]Fundación Pública Galega de Medicina Xenómica and Clinical University Hospital of Santiago de Compostela-SERGAS, Spain.
[4]Center for Biomedical Research on Rare DIseases (CIBERER) _ Institute of Health Carlos III, Spain
5-Keizo Asami Laboratory and Biological Sciences Graduate Program, Federal University of Pernambuco, Recife, Brazil.
6- Neuropsichiatry Department and Keizo Asami Laboratory, Federal University of Pernambuco,Recife, Brazil.
7-Human Genome Center, University of São Paulo, São Paulo, Brazil.
8 Department of Neurology, Hospital of Bellvitge, Barcelona, Spain.
9 Department of Neurology, Hospital of Figueres, Girona.
10 Department of Neurology, Monforte de Lemos Hospital-SERGAS, Lugo, Spain.
11 Belgrade Medical School, University of Belgrade, Belgrade, Serbia.
12 Neurology Clinic, University Clinical Center, Belgrade, Serbia.
13 Buriram hospital, Thailand

[*]Corresponding authors at:
Giovanni Coppola, MD
Semel Institute for Neuroscience and Human Behavior
1524 Gonda
695 Charles Young Drive South
90095 Los Angeles CA
gcoppola@ucla.edu
or Daniel H. Geschwind, MD PhD

Program in Neurogenetics, Department of Neurology
2306 Gonda
695 Charles Young Drive South
90095 Los Angeles CA
dhg@ucla.edu

**Mutations in *SLC20A2* are a Major Cause of Familial Idiopathic Basal Ganglia Calcification**

Sandy Chan Hsu, MS[1], Renee Sears, BA[1], Roberta R Lemos[5], Beatriz Quintáns[2,3], PhD, Alden Huang[1], Elizabeth Spiteri, PhD[1¶], Lisette Nevarez[1§],Zatz M[7], Vladimir Kostić[11], Milena Janković[11], Valerija Dobrić[12], Ivana Novaković[11],Peter Schofield, Kerrie Pierce, Jan Fullerton, Phillip Mitchell, Henry Brodaty, Jill Goldman, Melissa KMaisenbacher, Jennifer Mueller, Matthew Bower, MS, Martin PaucarArce, MD, Witoon Mitarnun[13], Agnes Linglart, Sheila Simpson, ZofiaMiedzybrodzka, PhD, Joanna Jen, MD PhD, Henry Paulson, MD PhD, Brent Fogel, MD PhD, Michele Yang[3],Serge Jaumà[8], Olga Carmona[9], Daniel García-Estevez[10], Catherine Mamah, MD[1§§], Joao RM de Oliveira[5,6], María-JesúsSobrido, MD PhD[2,4], Daniel H. Geschwind, MD PhD[1,3*], Giovanni Coppola, MD[1,3*]

[1]Program in Neurogenetics, Department of Neurology and [3]Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles CA
[2]Fundación Pública Galega de Medicina Xenómica and Clinical University Hospital of Santiago de Compostela-SERGAS, Spain.
[4]Center for Biomedical Research on Rare DIseases (CIBERER) _ Institute of Health Carlos III, Spain
5-Keizo Asami Laboratory and Biological Sciences Graduate Program, Federal University of Pernambuco, Recife, Brazil.
6- Neuropsichiatry Department and KeizoAsami Laboratory, Federal University of Pernambuco,Recife, Brazil.
7-Human Genome Center, Universityof São Paulo, São Paulo, Brazil.
8 Department of Neurology, Hospital of Bellvitge, Barcelona, Spain.
9 Department of Neurology, Hospital of Figueres, Girona.
10 Department of Neurology, Monforte de Lemos Hospital-SERGAS, Lugo, Spain.
11 Belgrade Medical School, University of Belgrade, Belgrade, Serbia.
12 Neurology Clinic, University Clinical Center, Belgrade, Serbia.
13 Buriram hospital, Thailand
Dr. Peter Schofield's affiliation :
Dr. Jill Goldman's affiliation : The Center for Parkinson's Disease & Related Disorders at Columbia University Medical Center
Melissa Maisenbacher's affiliation : Department of Pediatrics, University of Florida
Jennifer Mueller's affiliation : Department of Pediatrics, University of Florida
Matthew Bower's affiliation : University in Minnesota
Dr. Martin Arce's affiliation : Karolinska Institute
Dr. Agnes Linglart's affiliation :
Dr. Sheila Simpson's affiliation :
Dr. Zofia Miedzybrodzka's affiliation :
Dr. Henry Paulson's affiliation : Department of Neurology, University of Iowa
[*]Corresponding authors at:
Giovanni Coppola, MD
Semel Institute for Neuroscience and Human Behavior
1524 Gonda
695 Charles Young Drive South
90095 Los Angeles CA

gcoppola@ucla.edu

or

Daniel H. Geschwind, MD PhD
Program in Neurogenetics, Department of Neurology
2306 Gonda
695 Charles Young Drive South
90095 Los Angeles CA
dhg@ucla.edu

**ABSTRACT**
**Background:** Familial idiopathic basal ganglia calcification (IBGC) or Fahr's disease is a rare neurodegenerative disorder characterized by calcium deposits in the basal ganglia and other brain regions causing neuropsychiatric and motor symptoms. Familial IBGC is genetically heterogeneous and typically transmitted in an autosomal dominant fashion. We preformed a mutational analysis of *SLC20A2*, the first gene found to cause IBGC, to assess its genetic contribution to familial IBGC.
**Methods:** We recruited 218 subjects from 29 IBGC-affected families of varied ancestry and collected medical history, neurological exam, and head CT scans to characterize each patient's disease status. We screened our patient cohort for mutations in *SLC20A2* and used online databases and bioinformatic tools to assess the novelty and pathogenic potential of all identified variants.
**Findings:** 12 novel sequence variants (three nonsense variants, three deletions, three-splice site variants, and three missense variants), one rare synonymous variant and one previously reported mutation were identified in 13 families. In 5 families variants predicted to be deleterious co-segregated with disease. Three families showed non-segregation with clinical disease of such variants, but retrospective review of clinical and neuroimaging data strongly suggested previous mis-classification. Overall, mutations in *SLC20A2* account for as high as 41% of our familial IBGC cases.
**Interpretation:** Our screen in a large series expands the catalogue of *SLC20A2* mutations identified to date and demonstrates that mutations in *SLC20A2* are a major cause of familial IBGC. Non-perfect segregation patterns of predicted deleterious variants highlight the challenges of phenotypic assessment in this condition with highly variable clinical presentation.

## Introduction

Familial idiopathic basal ganglia calcification (IBGC) or Fahr's disease is an enigmatic neurodegenerative disorder characterized by calcium deposits in the basal ganglia and other brain regions in the absence of metabolic abnormalities or other causes of secondary calcification, such as infectious disease. Approximately 0.5 – 1.0% of CT scans in patients over age 50 exhibit incidental calcification of the basal ganglia that is sporadic and not familial.In contrast, familial IBGC is typically transmitted in an autosomal dominant fashion and is genetically heterogeneous. More than 30 families with segregating Mendelian forms of IBGC have been reported in the literature,

however its true prevalence remains unknown.[1]Clinical features include a variable combination of neuropsychiatric and motor symptoms including dystonia, parkinsonism, ataxia, psychosis, dementia, chorea, and frontal-subcortical cognitive dysfunction.[2,3]The variability in clinical presentation and penetrance, as well as the presence of phenocopies and relatively high prevalence of other causes of secondary calcifications have been significant confounding factorsin elucidating the genetic basis of familial IBGC.[4–6]

Efforts to ascertain a genetic location responsible for IBGC have resulted in the identification of three genetic loci through linkage analysis: IBGC1 on chromosome 14 (14q13), IBGC2 on chromosome 2 (2q37), and IBGC3on chromosome 8 (8p21.1-8q11.23).[7–9]Recently, Wang et al. reported the first causative gene linked to IBGC by identifying seven IBGC families with mutations in *SLC20A2*, a gene located in the IBGC3 region that encodes for type III sodium-dependent phosphate transporter 2 (PiT2).[10]Here, we present a mutational analysis of *SLC20A2* in 218 patients from 29 IBGC families of varied ancestry to further examine the genetic contribution of *SLC20A2* mutations in a large cohort of IBGC families. We identified 12 novel sequence variants predicted to be deleterious, one rare synonymous variant, and one previously known mutation in 13 of these families. Our findings show that mutations in *SLC20A2* are a major cause of familial IBGC and expand the catalogue of *SLC20A2* mutations identified to date.

## Patients and Methods

*Patient Recruitment and Assessment.* We identified 218 individuals belonging to 29 IBGC-affected families collected through the UCLA Medical Center from a number of collaborating institutions. Some of these families were included in previous clinical or genetic studies (Table 1). Informed consent was obtainedand the investigation was approved by the UCLA Institutional Review Board.Medical history and neurological examinations were performed in all probands and additional family members for most families. Serum calcium and parathormone levels were assayed in at least one proband from most families to exclude calcium dysregulation and other metabolic disorders that would cause brain calcifications unrelated to familial IBGC.

*Neuroimaging.* Head CT scans were performed as part of the diagnostic workup or reviewed for the presence of calcifications or other brain abnormalities. Subjects with CT scans positive for calcification were given an affected disease status, while CT-

negative patients >50 years who remained asymptomatic until their death were assigned an unaffected disease status. Subjects whose CT scans were negative but were under the age of 50, or whose CT scan results were not available, were classified as unknown.

*Molecular Genetics and Analytical Methods.* Blood samples were obtained from the participants and genomic DNA was extracted using standard methods. Using published primer pairs, we amplified all of the exonic and flanking intronic regions of *SLC20A2*by PCR in two CT-positive affected patients from each family.[10]The PCR reaction solution and touchdown PCR cycling conditionswere prepared and optimized using standard procedures.The final purified ampliconswere sequenced in both forward and reverse directions by Sanger sequencing on the ABI 3730 platform (Applied Biosystems) to produce chromatogram traces that were analyzed using the CodonCode software (CodonCode Corporation).

When variants were identified, all available family members in each family were screened using variant-specific primer pairs following the protocol described above. Online databases of human genetic variation were used to assess the novelty of the variants identified: the National Heart, Lung, and Blood Institute (NHLBI) exome variant server (http://evs.gs.washington.edu/EVS/, accessed July 2012), dbSNP135 as reported in the UCSC Genome Browser (http://genome.ucsc.edu/), and the 1000 Genomes Project (http://www.1000genomes.org, 20100804release 12 May 2012). The pathogenic potential of the identified variants was predicted using SIFT (Sorting intolerant from tolerant (http://sift.bii.a-star.edu.sg/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), and Human Splicing Finder software (http://www.umd.be/HSF/HSF.html, May 2009 release).[11–13]

**Results**

A total of 218 subjects from 29 families of various ancestries were included in the study. Major clinical features and CT findings of our IBGC family cohort are summarized in Table 1. At least one affected subject from each family exhibited movement, psychiatric and/or cognitive symptoms typical of familial IBGC. In five families, several asymptomatic individuals were also classified as affected because of the significant bilateral basal ganglial calcifications identified on CT scans.

Sequence analysis ofthe 29 family probands identified one previously published mutation, one rare synonymous variant, and 12 novel variants in *SLC20A2* in 13 families, including three nonsense variants, three deletions, three-splice site variants,

and three missense variants (Figure 1, Table 2, Supplementary Figure). None of the 12 novel variants had been reported in dbSNP 135, the 1000 Genomes database, the NHLBI Exome Sequencing Project, or in the previous study from Wang et al.[10]We did not identify *SLC20A2*variants or mutations in 16 of the 29 IBGC families screened. There was no clear correlation between age of onset, severity of disease symptoms, or any particular clinical phenotype in IBGC families with *SLC20A2* mutations compared to IBGC families without *SLC20A2* mutations.

To further explore pathogenicity we studied the 12 novel variants for segregation and predicted deleteriousness. Three nonsense variants were discovered:1)c.514A>T (family F22), introducing the stop codon in exon 4 (p.Lys172*); 2) c.760C>T (family F15), introducinga stop codon in exon 7 (p.Arg254*); 3) c.1652G>A (family F20) introducinga stop codon in exon 9 (p.Trp551*).Three deletions altering the protein reading frame were identified: 1) a base pair deletion(c.508delT) leading to a premature stop codon in exon 4 (p.Leu170*, family F1); 2) a two-base pair deletion (c.583_584delGT), predicted to substitute a leucine for a valine followed by a frame shift terminating after 61 aberrant amino acids (p.Val195Leufs*61, family F5); and 3) a four base-pair deletion (c.1828_1831delTCCC)which substitutes an alanine for a serine followed by a frameshift and premature termination after 17 amino acids (p.Ser610Alafs*17, F2 family).

Three of the variants identified were located at natural 5’ donor splice sites: 1) c.1523+1G>A (family F7) was located one base pair immediately flanking the 3’ end of exon 8; 2) c.1794+1G>A (family F9) and 3) c.1794+1G>C (familyF29), were both located one base pair immediately flanking the 3’ end of exon 10. The substitution of a guanine to an adenine or cytosine one nucleotide adjacent to the exon changes the highly critical GU dinucleotide essential for splicing and would most likely result in skipping of the affected exon equating to a large deletion in the final protein product. For the F7 family the splice site variant would result in the loss of exon 8, the largest exon in *SLC20A2*, while in the F9 and F29 families the splice site variant would result in the loss of exon 10. In both cases, exon exclusion is predicted to introduce an early stop codon (p.Gly312Valfs*8 and p.Ser570Argfs*30).

In the F7 family, the splicing variant (c.1523+1G>A) was in linkage disequilibrium with a missense variant in the coding region of exon 8 (c.1145G>A) substituting an arginine for a glutamine (p.Arg382Gln). Two additional novel missense variants were identified: 1) c.1506C>A (family F24) substituting a glutamine for a histidine at residue

502 (p.His502Gln), predicted to be a critical region for the transport function;[14]and 2) c.1703C>T (family F19) causing the change of a leucine to a proline at codon 568 (p.Pro568Leu). We also identified a previously known single base pair mutation (c.1802C>T) resulting in the amino acid substitution of a leucine for a serine at residue 601 (p.Ser601Leu) in family F23.[10]Currently, this is the only *SLC20A2* mutation that has been reported in more than one IBGC-affected family, and it would be interesting to know whether these two families share a common founder. Finally, we identified a rare synonymous sequence variant with an allele frequency of 0·0009 in the 1000 Genomes database (c.1101C>G, in the F18 family, p.Pro367Pro) of unknown pathogenic significance.

The pathogenic potential of the non-synonymous coding variants was analyzed using both PolyPhen-2 and SIFTand all were predicted to be damaging to the protein product in at least one out of the two prediction tools and several variants were predicted to be damaging in both applications (Table 2). In particular, the variants p.Leu170* and p.Val195leufs*61 were predicted to induce nonsense-mediated decay, a surveillance mechanism that would result in a degradation of the aberrant RNA product analogous to a complete deletion of one copy of the *SLC20A2* gene.[11]

Co-segregation analysis was performed in the families for which DNA was available for more than one affected subject (8 of 13). Five families (F7, F19, F9, F15 and F20) demonstrated perfectsegregation with disease status (Table 3). In contrast, two putatively affected subjects in F1, two affected subjects in F5, and one affected subject in the F2 family did not carry the *SLC20A2* variant found in all other respective affected family members. Review of CT scans revealed a strong contrast between the subjects who tested positive for mutations, presenting with clearly abundant and symmetrical calcification typical of IBGC-affected individuals, and those mutation negative, who presented only minimal calcifications, most likely consistent with phenocopies (Figure 3). For the remaining 5 families,only one affected subject was available in 4 families (F22, F23, F24, and F29), so we were not able to ascertain a segregation pattern, whereas in the family with the synonymous change (F18) one affected and one subject of unknown status shared the mutation. Excluding the three families with non-perfect segregation, five out of 23 families where a segregation analysis was possible or 22%, have segregating deleterious mutations in *SLC20A2*. Overall, considering the likelihood of phenocopies in families F2 and F5, and the predicted pathogenicity of the other

variants,*SLC20A2* variants and mutations may account for as high as 41%, or 12 out of 29, IBGC-affected families in our patient population.

**Discussion**

The recent identification of loss-of-function *SLC20A2* mutations in familial IBGC-affected patients finally advances the understanding of the molecular etiology of IBGC by establishing the first genetic location responsible for this disease.[10]Our systematic screen of 29 IBGC families identified one previously reported mutation, one rare synonymous variant,and 12 novel *SLC20A2* variants predicted to be deleterious, with at least five showing full segregation with disease status, indicating that mutations in *SLC20A2*are a major cause of familial IBGC. Furthermore, we identified *SLC20A2* variants in IBGC families of multiple ancestries, supporting the conclusion that *SLC20A2*mutations are linked to IBGC worldwide.

Nine out of the 14 mutations we identified are predicted to introduce a stop codon, pointing to haploinsufficiency as a causal mechanism for IBGC due to mutations in *SLC20A2*. Additionally, previous studies identifying the histidine at residue 502,[14] a position found to harbor a variant in family F24 (p.His502Gln), as critical for transport functionality in PiT2 highlights the loss of phosphate transport capacity as a major factor in the molecular etiology of IBGC. Also notable is that three of the four missense variants identified in our IBGC cohort are located within the ProDom domain (PD001131) shared by all PiT transporters (Figure 1).

Defining disease status in IBGC is complicated by several factors that have likely hampered identification of a clear genetic linkage signals: first, the broad variability of symptom manifestations, ranging from migraine and minor psychiatric symptoms to severe movement and cognitive disorders; second, the number of additional neurologic and systemic diseases that may present secondary brain calcifications; third, the common occurrence of age-related, idiopathic calcium deposits in the basal ganglia. While some IBGC family members with basal ganglia calcification are asymptomatic, others reporting neuropsychiatric or motor symptoms are CT-negative for calcifications. This poses the question as to whether the onset age of basal ganglia calcifications is variable in these patients, or whether their symptoms have a different etiology. The minimum age at which absence of calcifications at a CT scan excludes the disease remains unknown, contributing to ambiguity in identifying patients that harbor a pathogenic mutation but are asymptomatic and CT-negative at the time of data collection. Although CT status may not completely reflect disease status, both because it

can be normal in younger family members and the presence of unspecific calcifications in older individuals, it is currently the most reliable test for diagnosing IGBC.

We found a co-segregation of mutations with disease in five out of the eight families where a co-segregation analysis was possible (Table 3), consistent with the families reported by Wang et al.[10]We did not identify variant carriers who were not affected suggesting 100% sensitivity of the clinical/CT evaluation.In contrast, two out of 11 affected family members from the F1 family, two out of ten affected from the F5 family, and one out of the ten affected from the F2 family, had received affected disease status based upon clinical examination and/or CT scan, but did not carry predicted-deleterious *SLC20A2* sequence variants. Possible reasons for this finding include 1) incorrect clinical evaluation or CT scan analysis, and therefore suboptimal specificity of the clinical/CT-based diagnosis; CT calcifications observed in some of these patients are minor (Figure 3) and compatible with an incidental finding that appears in 0.5% - 1.0% of routine CT scans and is unrelated to familial IBGC[4–6]– for example, both of the F5 individuals who did not harbor the variant identified in the F5 family were asymptomatic and classified as affected based solely on CT scan; 2) non-causality of the identified sequence variants, unlikely given the predicted deleteriousness of the sequence variants (all microdeletions leading to frameshift causing nonsense mediated decay) and the cosegregation with disease in the vast majority of other family members; and possible technical factors include 3) false-negative mutation detection (non amplification or degradation of the mutated allele), unlikely since the mutation is detected in other family members; and 4) sample mix-up or labeling errors.

Importantly, the discovery of a novel and predicted deleterious *SLC20A2* variant in the F1 family, which we previously reported to have significant linkage to disease at the 14q13 (IBGC1) locus, suggests that the genetic mutation responsible for IBGC in this family was not on chromosome 14 but rather on chromosome 8 within *SLC20A2*. Notably, both individuals with discordant disease and genetic status contributing to the non-perfect segregation pattern observed in this family were also included in the initial cohort of 11 patients enrolled in the linkage mapping previously performed in this family. It is likely that the discordant disease status of these two individuals is due to clinical ascertainment and/or phenocopy, highlighting once again the importance of accurate phenotypic assessment when performing linkage studies, as linkage analysis relies heavily upon correct identification of affected individuals. The exclusion of

IBGC1 from linkage analysis studies in larger cohorts of distinct IBGC families has also demonstrated that IBGC1 is not a major genetic locus for this disease.[15]

The discovery of deleterious mutations in *SLC20A2* as a cause of familial IBGC greatly advances our understanding of this complex disease and will be crucial in the development of future treatments for IBGC patients as well as other conditions associated with brain calcification. Our assessment of the genetic contribution of *SLC20A2* mutations in our cohort of 218 familial IBGC patients demonstrates that as high as 41% (or 12 out of 29) of the collected families have predicted deleterious sequence variants or mutations in *SLC20A2*, strongly suggesting causality, and establishing *SLC20A2* as a key gene for familial IBGC. Furthermore, the identification of 12 novel variants – all predicted to be highly disruptive to protein function – broadens the spectrum of known *SLC20A2* mutations and adds to the genetic knowledge of this relatively unknown disease-causing gene.

More work is still needed to explain the variability in penetrance and expressivity within families. Identifying additional families in which genetic mutations correlate with symptom manifestation, such as the family we have described here, will provide valuable insight for understanding the molecular etiology responsible for the clinical heterogeneity observed in IBGC patients.

**Acknowledgements**

We would like to acknowledge and thank all of the participants and families for their valuable contribution to our study. This work was supported in part by CNPq (ConselhoNacional de DesenvolvimentoCientífico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior to… ), and FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco to….), NINDS (R01 … to D.H.G.).

References


1 Sobrido MJ, Hopfer S, Geschwind DH. Familial Idiopathic Basal Ganglia Calcification. In: Pagon R, Bird T, Dolan C, eds. GeneReviews [Internet]. Seattle WA, 2007.

2 Moskowitz M, Winickoff R, Heinz E. Familial calcification of the basal ganglions: a metabolic and genetic study. *New Engl J Med* 1971; **285**: 72–7.

3 Brodaty H, Mitchell P, Luscombe G, *et al.* Familial idiopathic basal ganglia calcification (Fahr's disease) without neurological, cognitive and psychiatric symptoms is not linked to the IBGC1 locus on chromosome 14q. *Hum Genet* 2002; **110**: 8–14.

4 Koller WC, Cochran JW, Klawans HL. Calcification of the basal ganglia: computerized tomography and clinical correlation. *Neurology* 1979; **29**: 328–33.

5 Harrington MG, Macpherson P, McIntosh WB, Allam BF, Bone I. The significance of the incidental finding of basal ganglia calcification on computed tomography. *Journal of neurology, neurosurgery, and psychiatry* 1981; **44**: 1168–70.

6 Förstl H, Krumm B, Eden S, Kohlmeyer K. What is the psychiatric significance of bilateral basal ganglia mineralization? *Biological psychiatry* 1991; **29**: 827–33.

7 Geschwind D, Loginov M, Stern J. Identification of a locus on chromosome 14q for idiopathic basal ganglia calcification (Fahr disease). *Am J Hum Genet* 1999; **65**: 764–72.

8 Volpato CBB, De Grandi A, Buffone E, *et al.* 2q37 as a susceptibility locus for idiopathic basal ganglia calcification (IBGC) in a large South Tyrolean family. *Journal of molecular neuroscience : MN* 2009; **39**: 346–53.

9 Dai X, Gao Y, Xu Z, *et al.* Identification of a novel genetic locus on chromosome 8p21.1-q11.23 for idiopathic basal ganglia calcification. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 2010; **153B**: 1305–10.

10 Wang C, Li Y, Shi L, *et al.*Mutations in SLC20A2 link familial idiopathic basal ganglia calcification with phosphate homeostasis. *Nat Genet* 2012. doi:10.1038/ng.1077.

11 Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**: 3812–4.

12 Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research* 2009; **37**: e67.

13 Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–9.

14 Bøttger P, Pedersen L. Mapping of the minimal inorganic phosphate transporting unit of human PiT2 suggests a structure universal to PiT-related proteins from all kingdoms of life. *BMC Biochemistry* 2011; **12**: 21.

15 Oliveira JR, Spiteri E, Sobrido MJ, *et al.* Genetic heterogeneity in familial idiopathic basal ganglia calcification (Fahr disease). *Neurology* 2004; **63**: 2165–7.

16 Kostić VS, Lukić-Ječmenica M, Novaković I, *et al.*Exclusion of linkage to chromosomes 14q, 2q37 and 8p21.1-q11.23 in a Serbian family with idiopathic basal ganglia calcification. *Journal of neurology* 2011; **258**: 1637–42.

LEGEND TO TABLES

**Table 1:** Main Clinical and Neuroimaging features of IBGC families.

**Table 2:** SLC20A2 variants and mutations identified in 13IBGC families.

**Table 3:** Cosegregation Analysis of Variants in *SLC20A2*. Numbers of subject in each category are reported by SLC20A2 variant status. Shaded cells highlight possible co-segregation mismatches.

LEGEND TO FIGURES

**Figure 1: Structure Model of PiT2 protein with the variant locations**
Red residues denote nonsense variants, orange residues denote missense variants, blue residue denote splice site variants, purple residues denote indels, and green residues denote synonymous variants.

**Figure 2:** DNA sequence chromatograms of the *SLC20A2* variants identified in IBGC-affected families

**Figure 3:** Brain CT images of non-segregating individuals and positive controls
(a) CT classified as positive for a F1 family member but tested negative for the variant; (b) CT positive F1 family member with the variant showing abundant calcifications; (c) CT classified as positive for F5 family member at age 75 but tested negative for the variant; (d) CT-positive F5 family member with the variant showing abundant calcifications.

# Tabelas

**Table 1: Main Clinical and Neuroimaging features of IBGC families**

| IBGC Family | Geographic Origin/Descent | N. of available members | Affected | Unaffected | Unknown | Clinical Symptoms | Diagnostic Workup | Calcifications | Reference |
|---|---|---|---|---|---|---|---|---|---|
| F1 | North-American | 31 | 11 | 7 | 13 | Parkinsonism, focal dystonia, tremor, and dysphagia | Serum calcium and PTH | +++ (BBG, globuspallidus, thalamus, white matter) | [7] |
| F2 | European | 30 | 10 | | | 8 individuals normal neurological, cognitive, psychiatric features; 2 individuals Dementia, Parkinsonism, Bipolar I | | Dense calcification in at least one area * | [3] |
| F3 | | 30 | 7 | 5 | 18 | Parkinsonism, dizziness, slurred speech, and balance disorder, | Serum calcium and PTH | | NP |
| F4 | North-American/Swedish | 17 | 10 | 3 | 4 | Dementia, chorea, slurred speech, palilalia, gait disturbance, 5 asymptomatic | Serum calcium and PTH | +++ (BBG, white matter, cerebellum) | [15] |
| F5 | North-American/ Irish | 16 | 10 | 2 | 4 | Parkinsonism and dystonia, 3 asymptomatic | Serum calcium and PTH | +++ (BBG, thalamus, white matter) | [15] |
| F6 | Serbian | 14 | 6 | | | Parkinsonism, severe gait disturbances with | | | [16] |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | freezing of gait, and dyskinesia; 2 asymptomatic | | | |
| F7 | | 12 | 6 | 0 | 6 | Cramps and headaches | Serum calcium and PTH | | NP |
| F8 | German | 10 | 6 | 1 | 3 | Dizziness, epilepsy, headaches | Serum calcium and PTH | +++ (BBG) | [15] |
| F9 | North-American/ Chinese | 9 | 6 | 0 | 3 | Dizziness, dementia, muscle spasms, and cramps; 2 asymptomatic | Serum calcium and PTH | +++ (BBG, cerebellum) | |
| F10 | | 6 | 3 | 0 | 3 | | Serum calcium and PTH | | NP |
| F11 | | 5 | 2 | 0 | 3 | | | | NP |
| F12 | Spanish | 5 | 2 | | | No cognitive or movement disorder, neurosensorial hearing loss, myopia, astigmatism, migraine headache, scoliosis, pescavus. | | +++ (BBG , thalamus, cerebellum, brainstem, cortico-subcortical) | NP |
| F13 | Swedish | 4 | 4 | 0 | 0 | | | | NP |
| F14 | | 4 | 3 | 0 | 1 | Dysarthria, micrographia, balance disorder | Serum calcium and PTH | | NP |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| F15 | | 3 | 2 | 0 | 1 | Depression and muscle spasms | Serum calcium and PTH | | NP |
| F16 | Portuguese | 3 | 3 | 0 | 0 | Migraine, vertigo, anxiety, depression, personality and behavioral problems, intellectual and language delay | | ++ (BBG, dentate nuclei calcification by fifth decade), + (earlier decades) | NP |
| F17 | Thai | 3 | 2 | 0 | 1 | | Serum calcium and PTH | | NP |
| F18 | Libyan | 2 | 1 | 0 | 1 | | Serum calcium and PTH | | NP |
| F19 | | 2 | 1 | 1 | 0 | Clawed hand and slurred speech | Serum calcium and PTH | | NP |
| F20 | | 2 | 2 | 0 | 0 | | Serum calcium and PTH | | NP |
| F21 | Spanish | 2 | 2 | 0 | 0 | Dopa-responsive parkinsonism, dysarthria, subcortical cognitive impairment, stroke | | +++ (BBG, cerebellum) | NP |
| F22 | | 1 | 1 | 0 | 0 | Dystonia, cramps, depression, insomnia, headaches, and muscle spasms | Serum calcium and PTH | +++ (BBG, dentate nucleus) | NP |
| F23 | | 1 | 1 | 0 | 0 | | Serum calcium and PTH | | NP |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| F24 | French | 1 | 1 | 0 | 0 | | Serum calcium and PTH | | NP |
| F25 | | 1 | 1 | 0 | 0 | Headaches, movement disorders | Serum calcium and PTH | | NP |
| F26 | Scottish | 1 | 1 | 0 | 0 | | Serum calcium and PTH | | NP |
| F27 | Spanish | 1 | 1 | 0 | 0 | Dementia, psychiatric disorder, parkinsonism, facial palsy, leukemia | | +++ (BBG, cerebellum) | NP |
| F28 | Spanish | 1 | 1 | 0 | 0 | Mild cognitive impairment, ataxia | | +++ (BBG, subcortical, cerebellum) | NP |
| F29 | European | 1 | 1 | 0 | 0 | | Serum calcium and PTH | | NP |

BBGC: Bilateral basal ganglia calcification; NP: unpublished; PTH: parathormone

**Table 2: 14 SLC20A2 variants identified in 13 IBGC families**

| N. | cDNA[a] | Amino-Acid[b] | Location | IBGC Family | Mutation Type | PolyPhen-2 Prediction | SIFT prediction | Human Splicing Finder |
|---|---|---|---|---|---|---|---|---|
| 1 | c.508delT | p.Leu170* | Exon 4 | F1 | Nonsense | Not available | Damaging and subject to nonsense mediated decay | |
| 2 | c.514A>T | p.Lys172* | Exon 4 | F22 | Nonsense | Not available | Damaging | |
| 3 | c.583_584delGT | p.Val195Leufs*61 | Exon 5 | F5 | Frameshift | Not available | Damaging and subject to nonsense mediated decay | |
| 4 | c.760C>T | p.Arg254* | Exon 7 | F15 | Nonsense | Not available | Damaging due to stop | |
| 5[d] | c.1101C>G | p.Pro367Pro | Exon 8 | F18 | Synonymous | Not available | Not available | |
| 6 | c.1145G>A | p.Arg382Gln | Exon 8 | F7 | Missense | Probably Damaging | Tolerated | |
| 7 | c.1506C>A | p.His502Gln | Exon 8 | F24 | Missense | Probably Damaging | Damaging | |
| 8 | c.1523+1G>A | p.Gly312Valfs*8 | IVS 8 | F7 | Splice site | | | Natural 5' Donor Site |
| 9 | c.1652G>A | p.Trp551* | Exon 9 | F20 | Nonsense | N/A | Damaging | |
| 10 | c.1703C>T | p.Pro568Leu | Exon 9 | F19 | Missense | Probably Damaging | Damaging | |
| 11 | c.1794+1G>A | p.Ser570Argfs*30 | IVS 10 | F9 | Splice site | | | Natural 5' Donor Site |
| 12 | c.1794+1G>C | p.Ser570Argfs*30 | IVS 10 | F29 | Splice site | | | Natural 5' Donor Site |
| 13[c] | c.1802C>T | p.Ser601Leu | Exon 11 | F23 | Missense | Probably Damaging | Damaging | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 14 | c.1828_1831delTCCC | p.Ser610Alafs*17 | Exon11 | F2 | Frameshift | N/A | Damaging (nonsense mediated decay not predicted) | |

[a]Numbering according to GenBank reference build NM_006749.3 starting at the translation initiation codon

[b]Numbering according to GenPept reference build NP_006740.1

[c]Previously reported by Wang et al. 2012[10]

[d]Rare variant with allele frequency of 0·0009 in the 1000 Genomes database

Table 3: Co-segregation Analysis of Variants in SLC20A2. Numbers of subject in each category are reported by SLC20A2 variant status.

Shaded cells highlight possible co-segregation mismatches.

| IBGC Family | Proband Variant | Family members with variant | | | Members without variant | | |
|---|---|---|---|---|---|---|---|
| | | Affected | Unaffected | Unknown | Affected | Unaffected | Unknown |
| F1 | c.508delT | 9 | | 2 | 2 | 8 | 10 |
| F2 | c.1828_1831delTCCC | 9 | | 1 | 1 | 6 | 13 |
| F5 | c.583_584delGT | 8 | | | 2 | 2 | 4 |
| F7 | c.1523+1G>A<br>c.1145G>A | 6 | | 1 | | | 5 |
| F9 | c.1794+1G>A | 6 | | 1 | | | 2 |
| F15 | c.760C>T | 2 | | | | | 1 |
| F19 | c.1703C>T | 1 | | | | 1 | |
| F18 | c.1101C>G | 1 | | 1 | | | |
| F20 | c.1652G>A | 2 | | | | | |
| F22 | c.514A>T | 1 | | | | | |
| F23 | c.1802C>T | 1 | | | | | |
| F24 | c.1506C>A | 1 | | | | | |
| F29 | c.1794+1G>C | 1 | | | | | |

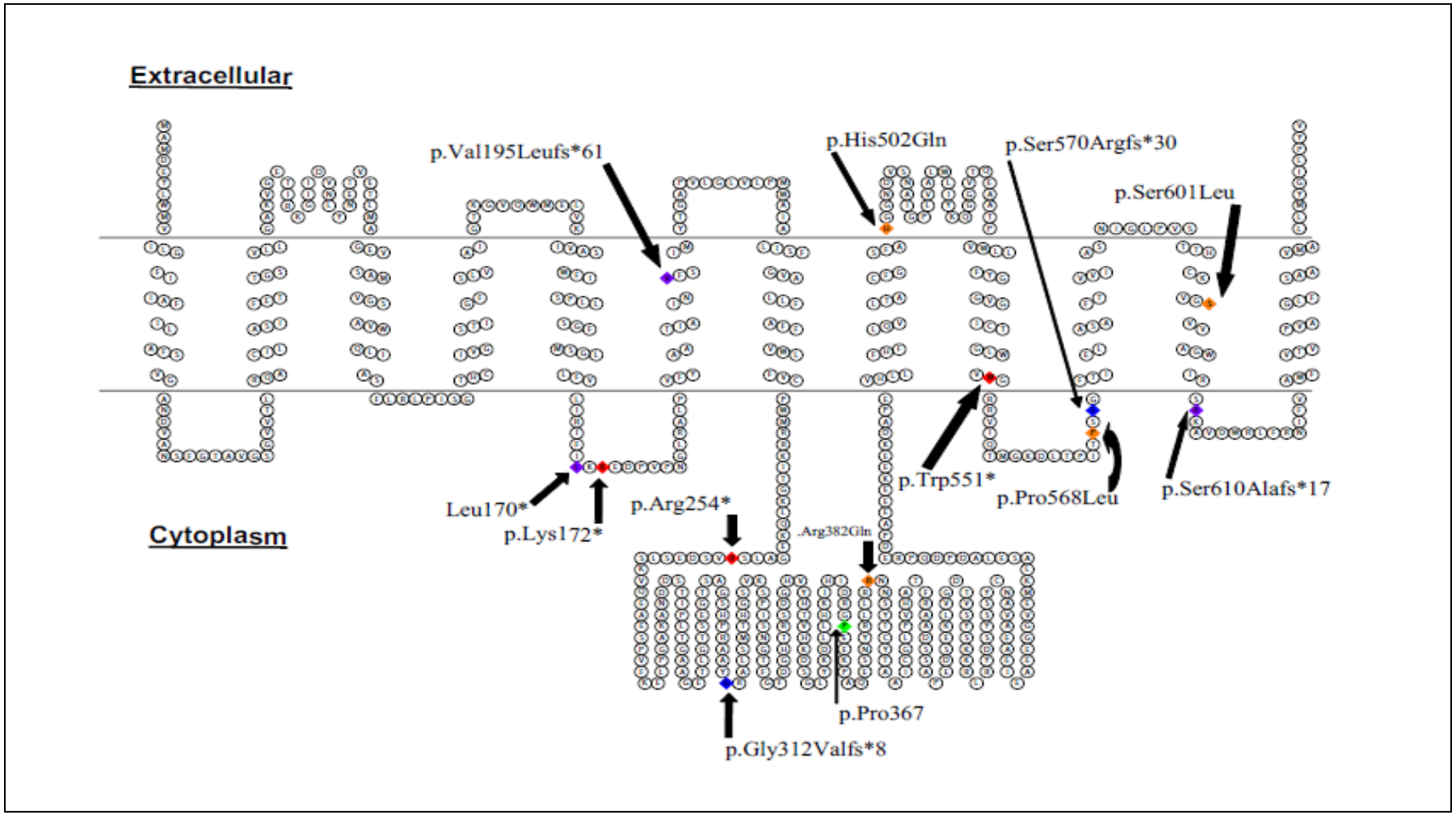



**Figure 1:** Structure Model of PiT2 protein with the variant locations. Red residues denote nonsense variants, orange residues denote missense variants, blue residue denote splice site variants, purple residues denote indels, and green residues denote synonymous variants.

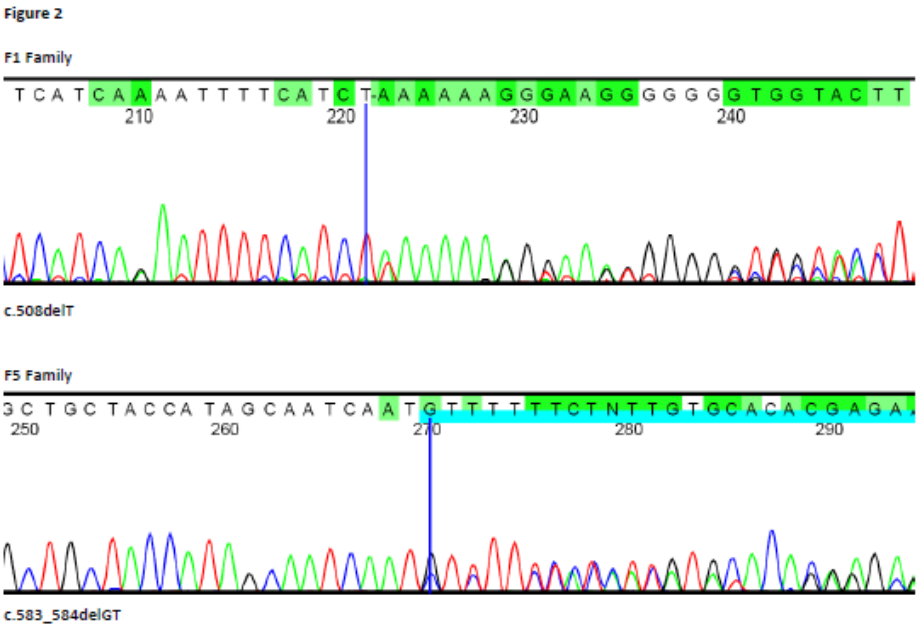



**Figure 2:** DNA sequence chromatograms of the *SLC20A2* variants identified in IBGC-affected families

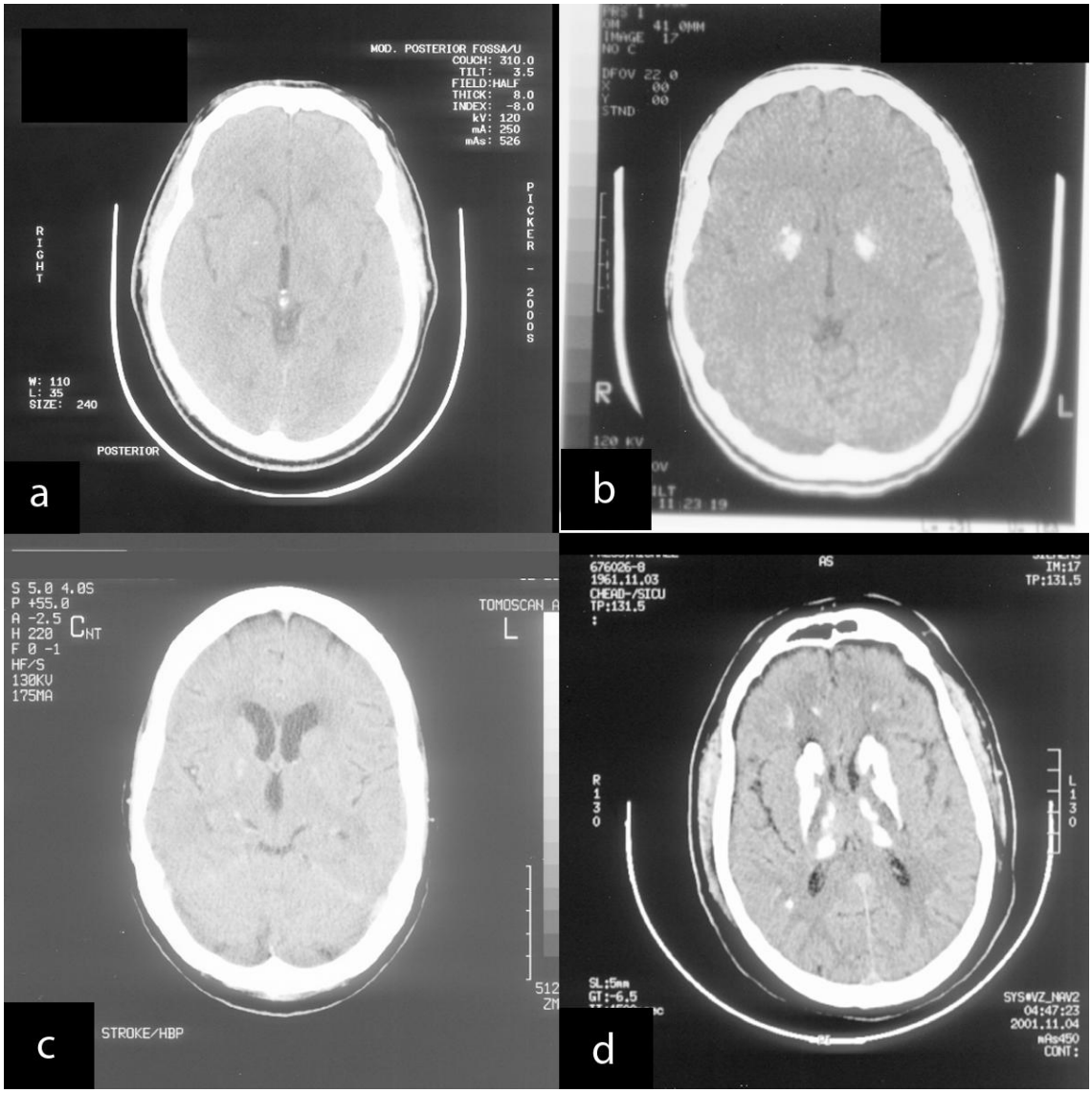



**Figure 3:** Brain CT images of non-segregating individuals and positive controls
(a) CT classified as positive for a F1 family member but tested negative for the variant; (b) CT positive F1 family member with the variant showing abundant calcifications; (c) CT classified as positive for F5 family member at age 75 but tested negative for the variant; (d) CT-positive F5 family member with the variant showing abundant calcifications.

# Capítulo 5

# Artigo aceito para publicação

## Reporting a new mutation at the SLC20a2 gene in a Brazilian family with idiopatic basal ganglia calcification.

Corresponding author:

João Ricardo Mendes de Oliveira , Oliveira JRM

Department of Neuropsychiatry, Federal University of Pernambuco (UFPE), zip code: 50670-901 Recife, PE, Brazil

e-mail: joao.ricardo@ufpe.br

Phone: 55-81-87819856 Fax: 55-81-2126-8484

full names of the co-authors: Lemos RR[1], Oliveira MF[2], Oliveira JRM[1,3]

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

2- Neurosurgery residency program. Hospital do Servidor Público Estadual do Estado de São Paulo, São Paulo – SP, Brazil

3- Neuropsychiatric Department - Federal University of Pernambuco (UFPE), Recife, Pernambuco-Brazil

EUROPEAN JOURNAL OF NEUROLOGY
THE OFFICIAL JOURNAL OF THE EUROPEAN FEDERATION OF NEUROLOGICAL SOCIETIES

# Reporting a new mutation at the SLC20a2 gene in familial idiopatic basal ganglia calcification.



| | |
|---|---|
| Journal: | *European Journal of Neurology* |
| Manuscript ID: | EJoN-12-0932 |
| Wiley - Manuscript type: | Letters to the Editor |
| Date Submitted by the Author: | 05-Sep-2012 |
| Complete List of Authors: | Lemos, Roberta; Universidade Federal de Penambuco, Neuropsychiatry<br>Oliveira, Matheus; Hospital do Servidor Público Estadual de São Paulo, Neurosurgery<br>Oliveira, João; Universidade Federal de Penambuco, Neuropsychiatry |
| Keywords: | Ataxia and gait disorders < Movement disorders < NEUROLOGICAL DISORDERS, Parkinson's disease < Movement disorders < NEUROLOGICAL DISORDERS, Psychosis < Psychiatric disorders < NEUROLOGICAL DISORDERS, Depression < Psychiatric disorders < NEUROLOGICAL DISORDERS, Adult < PATIENT CATEGORIES |
| | |

# Reporting a new mutation at the SLC20a2 gene in familial idiopatic basal ganglia calcification.

Corresponding author:

João Ricardo Mendes de Oliveira , Oliveira JRM

Department of Neuropsychiatry, Federal University of Pernambuco (UFPE), zip code: 50670-901 Recife, PE, Brazil

e-mail: joao.ricardo@ufpe.br

Phone: 55-81-87819856 Fax: 55-81-2126-8484

full names of the co-authors: Lemos RR[1], Oliveira MF[2], Oliveira JRM[1,3]

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

2- Neurosurgery residency program. Hospital do Servidor Público Estadual do Estado de São Paulo, São Paulo – SP, Brazil

3- Neuropsychiatric Department - Federal University of Pernambuco (UFPE), Recife, Pernambuco-Brazil

Abstract

Mutations at the SLC2A2 gene were recently linked to idiopathic basal ganglia calcification (IBGC) in families from China, Spain and Brazil. This finding associated IBGC with the unbalance of inorganic phosphate homeostasis. Now we report a new missense mutation at this gene, at the exon 8, position c. 1483G>A (A495T) and predicted as pathogenic by SIFT, I-Mutant2.0 and Polyphen. There is no report of such variation at the 1000 Genome database or the ESP6500 data set from the National Heart, Lung, and Blood Institute Exome Sequencing Project. This kindred, with Caucasian background, present a remarkable clinical heterogeneity with the youngest affected presenting more symptoms, and more calcifications, when compared to his mother, also affected but with milder symptoms. The rest of the family was not available for additional analysis. Idiopathic brain calcinosis presents a heterogeneous clinical profile, including a variable combination of motor and cognitive symptoms. The study of new families with this condition will be crucial to identify other genes and our finding reinforces the relevance of the SLC20A2 gene to the etiopathogeny of IBGC, pointing the exon 8 at the one, so far, with more mutations report.

Keywords: SLC20A2, basal ganglia calcification, Fahr's Disease, parkinsonism, clinical heterogeneity

**Dear Editor,**

Mutations at the SLC2A2 gene, at the chromosome 8, were recently linked to familial idiopathic basal ganglia calcification in families from China, Spain and Brazil [1].

Due to the increasing use of neuroimaging procedures, brain calcifications are visualized more often than ever before. Patients burdening these calcifications, mainly in basal ganglia but occasionally also in cerebellum and white matter, present a wide variety of symptoms such as parkinsonism, headaches, psychosis, dementia and mood symptoms [2].

We have collected clinical history and neuroimages from two available subjects from a Brazilian family with idiopathic brain calcinosis and Caucasian background. The rest of the family was not available for additional analysis.

These results are part of a study approved by the ethics committee from the Federal University of Pernambuco (CAEE-0296.0172.000-08).

The figure 1 shows the pedigree chart, an eletropherogram displaying the misense mutation and 3D models of brain calcifications in each affected, using 3D-Doctor software (Able Software, Lexington, MA, USA).

The clinical manifestation of I-2, a 84 years old woman, is mild depression and parkinsonism of late manifestation, despite the presence of large calcifications (10,85 $cm^3$) in basal ganglia and cerebellum. The II-1 subject presents calcifications in basal ganglia, pineal and plexus choroid (20,73 $cm^3$). The rest of the family was not available for a thorough analysis

II-1 is a 43-year old man, one of the seven children from I-2, with rapid progression of parkinsonism features in the late third decade of life. Until then, he was a fully

functional person with regular employment. In the last 5 years, a progressive presentation of general bradykinesia, rigidity and paresis in right arm has developed. Recently, cognitive and speech disorders were added, characterized by dysarthria, mild cognitive impairment and moderate depressive state. No other cause was disclosed, such as stoke, demyelinating conditions, autoimmune processes or metabolic disorder.

During the screening of mutations at the SLC20a2 gene, using automated sequencing in a MegaBACE 1000 (Sunnyvale, CA), we identified a new missense mutation at the exon 8, position c. 1483G>A (A495T), predicted to be pathogenic by SIFT, I-Mutant2.0 and Polyphen analysis.

This is a novel mutation with no previous report at the 1000 Genomes Project databases, at the ESP6500 data set from the National Heart, Lung, and Blood Institute (consisting of 4300 exomes from European-Americans and 2203 from African-American individuals) or in other 4 Brazilian families excluded during the same screening.

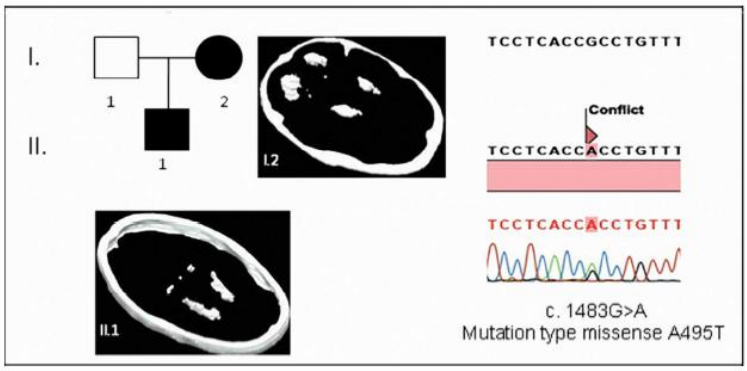
This finding reinforces the relevance of the SLC20A2 gene to the etiopathogeny of IBGC and point the exon 8 as a region, so far, with more mutations reported.

Ongoing analysis of genome wide scans in additional families showed new candidate regions at chromosomes 2, 7, 9 and 14, indicating genetic heterogeneity. The study of new families with this condition will be crucial to identify other genes and to reinforces the relevance of the SLC20A2 gene to the etiopathogeny of IBGC [3,4].

**URLs. 1000 Genomes Project**, http://www.1000genomes.org/; **NHLBI Exome Sequencing Project** (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/; **Sift**, http://sift.jcvi.org/www/SIFT_enst_submit.html; **Polyphen**: http://genetics.bwh.harvard.edu/pph2/; **I-Mutant2.0**: http://folding.uib.es/i-mutant/i-mutant2.0.html

**Acknowledgements:**

We are greatly indebted with Dr Joana Braga for positional analysis. This study was supported by grants from the following Brazilian funding agencies and academic bureaus: CNPq, FACEPE, PROPESQ-UFPE, PIBIC-UFPE and CAPES. Dr. João Oliveira holds a CNPq Fellowship (Brazil).




Presentation of heredogram with subjects I-2 and II-1, with respective Skull CTs revealing basal ganglia calcifications. Additionally, the eletropherogram of subject II-1 pointing new mutation.
69x33mm (300 x 300 DPI)



Errata:

Onde se lê: c.1483C>A, leia-se: c.2140C>A.

## 6.0 Conclusões

As análises desenvolvidas nesta tese puderam contribuir para o início da compreensão das bases genéticas da IBGC, bem como para o entendimento de metodologias integradas para identificação de fatores de risco genético.

Para DA foi possível validar em partes o *pipeline* de bioinformática proposto, uma vez que variações do tipo SNPs foram detectas, porém pequenos INDELs de 1 a 10pb, não foram encontrados. Alternativamente, as abordagens usadas nas plataformas NGS poderão auxiliar na detecção desse tipo de variação.

A ausência da mutação Pro521Ala no gene MGEA6 (lócus IBGC1), nas duas famílias Brasileiras afetadas e no grupo controle, demonstra o quanto essa mutação é rara, esses resultados contribuíram para a diminuição do MAF dessa variante.

Evidências genéticas indicam que as mutações no gene SLC20A2 lócus IBGC3, são responsáveis por mais de 40% dos casos clínicos investigados. Um novo relato de outra família Brasileira com mutação no gene SLC20A2, reforça a importância desse gene como sendo, atualmente a principal causa genética da IBGC.

Projetos de exomas estão sendo desenvolvidos para as famílias excluídas do gene SLC20A2. Essas análises serão futuramente compartilhadas entre os grupos de interesse, a fim de acelerar a identificação de novos genes candidatos.

As futuras descobertas irão refletir no diagnóstico e tratamento dessas patologias, através do desenvolvimento de estratégias de menor custo, rápida execução e eficaz resposta terapêutica.

# Anexos

**Artigo de outros temas**

1. 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. , v.xxx, p.xxx - , 2012.


Hypertension Research

Access
To read this article in full you may need to log in, make a payment or gain access through a site license (see right).

nature.com > Journal home > Table of Contents

Original Article

*Hypertension Research* **35**, 733-738 (July 2012) | doi:10.1038/hr.2012.14

Exaggerated blood pressure response during exercise treadmill testing: functional and hemodynamic features, and risk factors

Sandro G de Lima, Maria de F P M de Albuquerque, João R M de Oliveira, Constância F J Ayres, José E G da Cunha, Danyllo F de Oliveira, Roberta R de Lemos, Manuela B R de Souza and Odwaldo B e Silva

ARTICLE TOOLS
Send to a friend
Export citation
Rights and permissions
Order commercial reprints
Bookmark in Connotea

SEARCH PUBMED FOR

**2.** LEMOS, R. R., LIMA, S. G., Cunha, J. E. Gomes, OLIVEIRA, D. F., Souza, M. B. Rodrigues, AYRES, C. F. J., ALBUQUERQUE, M. F. P. M., OLIVEIRA, J. R. M.

Revising the M235T Polymorphism Position for the AGT Gene and Reporting a Modifying Variant in the Brazilian Population with Potential Cardiac and Neural Impact. Journal of Molecular Neuroscience. , v.000, p.000 - , 2012.

**Resumos de Congressos**

SOUZA, K. T., LEMOS, R.R., OLIVEIRA, D.F, OLIVEIRA, J.R.M. Descrição da primeira família Brasileira com "Doença de Fahr" (Calcificação Idiopáticas em núcleos da Base do cérebro) ligada a uma mutação no gene SLC20A2. In: XXV Congresso Brasileiro de Neurologia, Goiana, 2012.

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 M235Tpolymorphism 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.
**Manipulação do DNA: 4 décadas rompendo fronteiras**. , 2011. v.000. p.000 - 000

LEMOS, R.R., SOUZA, M.B.R, OLIVEIRA, D. F., PIMENTEL, L. F., OLIVEIRA, J. R. M. Troubleshooting the search for potential pathogenic small INDELs (1-10pb) in neuropsychiatric disorders In: 57º Congresso Brasileiro de Genética, 2011, Águas de Lindóia. **Manipulação do DNA: 4 décadas rompendo fronteiras**. , 2011. v.000. p.000 - 000

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: 65º 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: 65º Congresso Brasileiro de Cardiologia, 2010, Belo Horizonte.
**Arquivos Brasileiros de Cardiologia**. , 2010. v.95. p.47 - 47 Página gerada pelo Sistema Currículo Lattes em 24/06/2012 as 19:42:40 Página 5 de 8

LEMOS, R. R., OLIVEIRA, M. F., OLIVEIRA, D. F., ZATZ, M., OLIVEIRA, J. R. M. NEUROIMAGING AND MOLECULAR ANALYSIS IN FAMILIAL IDIOPATHIC BASAL GANGLIA CALCIFICATION ("FAHR DISEASE") In: XXXIV Congresso Anual da Sociedade Brasileira de Neurociências e Comportamento SBNEC-2010, 2010, Caxambú. **Sociedade Brasileira de Neurociências e Comportamento - SBNeC 2010**. , 2010.

OLIVEIRA, J. R. M., OLIVEIRA, M. F., KUHNI, R., LEMOS, R. R., OLIVEIRA, D. F. Neuro-imaging genetics studies in basal ganglia calcification as a model to understand brain resilience In: The Israel Society for Neuroscience 19th Annual Meeting, 2010, Eilat. **Israel Society for Neuroscience**. , 2010. v.0. p.0 - 0

OLIVEIRA, J. R. M., LEMOS, R. R., SOUZA, M.B.R, CUNHA, J.E.G., MARQUES, E. T., LIMA FILHO, J. L. Em busca de novos fatores de risco genético para transtornos neuropsiquiátricos em bancos de dados de expressão In: 55º Congresso Brasileiro de Genética, 2009, Águas de Lindóia.
**55º Congresso Brasileiro de Genético**. , 2009.

LEMOS, R.R., OLIVEIRA, M. F., OLIVEIRA, J. R. M. Neuroimaging and Molecular Analysis of a Brazilian family with idiopathic basal ganglia calcification (Fahr Disease) In: SINATER, 2009, Recife. **II Simpósio Nacional em Diagnóstico e Terapêutica Experimental, V Jornada Científica do LIKA e II Fórum Brasileiro de Genética em Neuropsiquiatria**. , 2009.

OLIVEIRA, D. F., LEMOS, R. R., OLIVEIRA, J. R. M.
Populational and Computational analysis of the MGEA6 P521A Polymorphism in Fahr's Disease In: SINATER, 2009, Recife.
**II Simpósio Nacional em Diagnóstico e Terapêutica Experimental, V Jornada Científica do LIKA e II Fórum Brasileiro de Genética em Neuropsiquiatria**. , 2009.

SOUZA, M.B.R, Lemos, R.R., CUNHA, J.E.G., LIMA FILHO, J. L., OLIVEIRA, J. R. M. Searching for new genetic risk factorsfor neuropsychiatric disordrs in expression databases In: SINATER, 2009, Recife.
**II Simpósio Nacional em Diagnóstico e Terapêutica Experimental, V Jornada Científica do LIKA e II Fórum Brasileiro de Genética em Neuropsiquiatria**. , 2009.

## Comprovantes de aceites para publicação

Ref.: Ms. No. NEUROGEN-D-12-00109
Mutations in SLC20A2 are a Major Cause of Familial Idiopathic Basal Ganglia Calcification
Neurogenetics

Dear Dr Coppola,

I am pleased to tell you that your work has been accepted for publication in Neurogenetics.

It was accepted on November 6, 2012, and you should receive the galleys from the publisher shortly.

Thank you for submitting your work to this journal.

With kind regards

M Graeber
Editor, Neurogenetics

---

**De:** ENE@wiley.com
**Para:** "joao ricardo" <joao.ricardo@ufpe.br>
**Enviadas:** Quinta-feira, 25 de Outubro de 2012 13:41:45
**Assunto:** Your article to be published in European Journal of Neurology - ENE-12044

Dear Author,

Journal: European Journal of Neurology

Article title: "Reporting a new mutation at the SLC20a2 gene in familial idiopatic basal ganglia calcification."

Your article has been received by our production department. In order to get access, you must register at our Wiley-Blackwell Author Services website.

By visiting Author Services and registering now, you will also be able to enjoy the following additional benefits:

- Track the progress of your article through the production process
- Receive publication status alerts
- Enable free access to your published article online for all co-authors and up to 10 nominated colleagues
- Enjoy a 25% discount on Wiley books

To register at Wiley-Blackwell Author Services and have your article automatically added to your account, go to:

http://authorservices.wiley.com/bauthor/addart.asp?articleid=1076705%2D1000310

If you have already registered with Author Services simply select the URL above and log in using your e-mail address and password. Use the "Add Article" feature to add this article to your profile; the Unique ID for this article is: 1076705-1000310. Your username is the e-mail you used to register.

WILEY-BLACKWELL

| Wiley Online Library | Home

# Author Services

**Author Services Menu**

- Home

**Journal Authors**

- Journal Authors Home
- Register
- My Publications
- Add New Article
- Amend My Details
- My Favorite Journals
- Find a Journal
- Editorial Policies
- Author Resources
- Author Rights and Benefits
- FAQs

**Book Authors**

- Book Authors Home

**My Publications**

**Dr Matheus Oliveira (mafernoliv@yahoo.com.br)**

Add a new article | Amend my details | Sign out

**Your Articles**


**European Journal of Neurology**

"Reporting a new mutation at the SLC20a2 gene in familial idiopatic basal ganglia calcification."

**Article Status** - View stage descriptions

1 2 3 4

**Accepted article received**
25 Oct 2012

Production History ||

Make my article OnlineOpen

Add My Colleagues

Add My Co-Authors

E-mail Journal Production Editor