Carlos Eduardo Malvezzi Kawamata

**IDENTIFICAÇÃO DE POSSÍVEIS GENES
RELACIONADOS COM A INFECÇÃO POR
Trypanosoma cruzi NO HOSPEDEIRO**

Tese apresentada ao Departamento de
Parasitologia do Instituto de Ciências
Biomédicas da Universidade de São Paulo, para
a obtenção do título de Doutor em Ciências.

São Paulo

2011

Carlos Eduardo Malvezzi Kawamata

**IDENTIFICAÇÃO DE POSSÍVEIS GENES
RELACIONADOS COM A INFECÇÃO POR
Trypanosoma cruzi NO HOSPEDEIRO**

Tese apresentada ao Departamento de Parasitologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para a obtenção do título de Doutor em Ciências.

Área de Concentração: Biologia da Relação Patógeno-Hospedeiro.

Orientador: Prof. Dr. Henrique Krieger.

Versão Original

São Paulo

2011

DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e Informação Biomédica do
Instituto de Ciências Biomédicas da Universidade de São Paulo

© reprodução total

Kawamata, Carlos Eduardo Malvezzi.

Identificação de possíveis genes relacionados com a infecção por *Trypanosoma cruzi* no hospedeiro humano / Carlos Eduardo Malvezzi Kawamata. -- São Paulo, 2011.

Orientador: Prof. Dr. Henrique Krieger.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Ciências Biomédicas. Departamento de Parasitologia. Área de concentração: Biologia da Relação Patógeno-Hospedeiro. Linha de pesquisa: Epidemiologia genética.

Versão do título para o inglês: Identification of possible genes related to *Trypanosoma cruzi* infection in the human host

Descritores: 1. Doença de chagas 2. Epidemiologia genética 3. Genechip I. Krieger, Prof. Dr. Henrique II. Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Biologia da Relação Patógeno-Hospedeiro III. Título.

ICB/SBIB0200/2011

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Carlos Eduardo Malvezzi Kawamata.

Título da Tese: Identificação de possíveis genes relacionados com a infecção por *Trypanosoma cruzi* no hospedeiro humano.

Orientador(a): Prof. Dr. Henrique Krieger.

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão pública realizada a/...../....., considerou

() Aprovado(a)

() Reprovado(a)

Examinador(a): Assinatura:
Nome completo:
Instituição:

Examinador(a): Assinatura:
Nome completo:
Instituição:

Examinador(a): Assinatura:
Nome completo:
Instituição:

Examinador(a): Assinatura:
Nome completo:
Instituição:

Presidente: Assinatura:
Nome completo:
Instituição:

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 - cep. 05508-000 São Paulo, SP - Brasil
Telefone : (55) (11) 3091.7733 - telefax : (55) (11) 3091-7438
e-mail: cep@icb.usp.br

São Paulo, 17 de fevereiro de 2006.

PARECER 687/CEP

Prezado Senhor,

Atendendo sua solicitação, a *Comissão de Ética em Pesquisas com Seres Humanos do ICB*, analisou o projeto de sua responsabilidade intitulado: "*Epidemiologia genética da infecção por Trypanosoma cruzi*".

Informo a V.Sa. que, após análise, o referido projeto foi **aprovado por esta Comissão.**

Esclareço a V.Sa. que dentro de 12 meses, relatório do referido projeto deverá ser encaminhado à Secretaria deste CEP.

Atenciosamente,

Prof. Dr. LUIZ VICENTE RIZZO
Coordenador da Comissão de Ética em
Pesquisas com Seres Humanos - ICB/USP

Ilmo. Sr.
PROF. DR. HENRIQUE KRIEGER
Departamento de Parasitologia
Instituto de Ciências Biomédicas - USP

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 - cep. 05508-000 São Paulo, SP - Brasil
Telefone: (55) (011) 3091.7733 - telefax: (55) (011) 3091-7438
e-mail: cep@icb.usp.br

Decl. CEP SH.007/2006

DECLARAÇÃO

Em adendo ao **Parecer 687/CEP**, de 17 de fevereiro de 2006, informo que o projeto desenvolvido pelo aluno **CARLOS EDUARDO MALVEZZI KAWAMATA**, *está inserido no projeto: "Epidemiologia genética da infecção por Trypanosoma cruzi"*, sob a responsabilidade do **Prof. Dr. Henrique Krieger**.

São Paulo, 17 de fevereiro de 2006

Atenciosamente,

Prof. Dr. LUIZ VICENTE RIZZO
Coordenador da Comissão de Ética em
Pesquisas com Seres Humanos - ICB /USP

Comissão de Ética em Pesquisa com Seres Humanos do Instituto de Ciências Biomédicas - USP
Aprovada pela Comissão Nacional de Ética em Pesquisa - CONEP, em 10 de fevereiro de 1998.

Aos meus pais pelo apoio e
companheirismo em todos os
momentos da minha vida.

AGRADECIMENTOS

Gostaria de agradecer ao grande amigo e orientador professor Doutor Henrique Krieger.

A toda a equipe que participou do projeto e da coleta dos dados da Hospedaria de Imigrantes do Departamento de Imigração e Colonização do Estado de São Paulo, em particular a Dra. Mary Furlan Feitosa.

Aos grandes amigos Dr. Ricardo de Godoi Matos Ferreira e Dr. Leandro Maza Garrido, que me apoiaram bastante.

A todos os amigos do laboratório de Epidemiologia Genética que estiveram presentes durante o meu curso de pós-graduação.

Ao Departamento de Parasitologia do Instituto de Ciências Biomédicas da Universidade de São Paulo.

Ao Instituto Nacional de Genética Médica Populacional (INAGEMP).

Ao CNPq, à FAPESP e à CAPES pelo apoio financeiro.

LISTA DE TABELAS

Tabela 1 – Comparação entre as duas fontes de lâminas de imunofluorescência	39
Tabela 2 - Concentração do DNA de cada indivíduo	51
Tabela 3 – 18 SNPs mais significativos do teste de associação	53
Tabela 4 – Valores de P para os 18 SNPs após a correção de Bonferroni.....	54
Tabela 5 – Mistura racial dos indivíduos do Grupo Caso.	55
Tabela 6 – Mistura racial dos indivíduos do Grupo Controle.	55
Tabela 7 – Tabela dos testes de equilíbrio de Hardy-Weinberg para os 18 SNPs.	56

LISTA DE FIGURAS

Figura 1 – Ciclo da doença de Chagas.....	27
Figura 2 – Mapa da área endêmica da doença de Chagas.....	29
Figura 3 – Memorial do Imigrante.....	34
Figura 4 – Foto gel de verificação 1	49
Figura 5 – Foto gel de verificação 2	50
Figura 6 – Gráfico do estudo de associação de 250 mil SNPs.....	52
Figura 7 – Ligação dois-pontos em todos os cromossomos com espaçamento de 50cM.....	60
Figura 8 – Ligação dois-pontos em todos os cromossomos com espaçamento de 25cM....	61
Figura 9 – Ligação multi-ponto em todos os cromossomos com espaçamento de 50cM.....	62
Figura 10 – Ligação multi-ponto no cromossomo 18 com espaçamento de 20cM.	63
Figura 11 – Ligação multi-ponto no cromossomo 10 com espaçamento de 20cM.	64

LISTA DE ANEXOS

Anexo 1 – Análise de regressão múltipla escalonada	83
Anexo 2 – Análise de correlação	83
Anexo 3 – Análises de componentes de variância	83
Anexo 4 – Análise de segregação complexa	84
Anexo 5 – Protocolo de genotipagem Affymetrix	85

RESUMO

KAWAMATA, C. E. M. **Identificação de possíveis genes relacionados com a infecção por *Trypanosoma cruzi* no hospedeiro humano.** 2011. 125p. Tese (Doutorado em Biologia da Relação Patógeno-Hospedeiro). Instituto de Ciências Biomédicas, Universidade de São Paulo, 2011.

A doença de Chagas é causada pelo protozoário *Trypanosoma cruzi* e atinge cerca de 12 milhões de pessoas no continente americano, a forma clássica de transmissão ocorre por intermédio do inseto vetor da subfamília Triatominae, popularmente chamado de barbeiro. Em um trabalho anterior, foi realizada uma análise de segregação complexa que indicou a presença de um gene principal, com um componente multifatorial influenciando a predisposição à infecção por *Trypanosoma cruzi*. A população é composta por 4697 indivíduos pertencentes a 886 famílias vindas do Nordeste do Brasil e tiveram os dados e amostras de sangue e saliva coletados entre 1969 e 1970. No presente estudo foi utilizada uma amostra de 69 indivíduos, sendo 18 positivos para a infecção por *Trypanosoma cruzi* e 51 negativos, distribuídos em 14 famílias. Os indivíduos tiveram seu DNA extraído e genotipado utilizando microarranjos de DNA de 260 K SNPs (GeneChip Mapping Affymetrix). Testes de associação mostraram significância entre a infecção por *T. cruzi* e o SNP rs17469997 do cromossomo 10, com $P=0,015$ após a correção de Bonferroni. Para validar estes inéditos resultados, análises de ligação multi-ponto foi feita com o programa GeneHunter (KRUGLYAK et al., 1996) e ligação dois-pontos com o programa SuperLink (FISHELSON e GEIGER, 2002), mas ambas não apresentaram resultados significativos, devido ao pequeno número de famílias informativas.

Palavras-chave: doença de Chagas. epidemiologia genética. genechip.

ABSTRACT

KAWAMATA, C. E. M. **Identification of possible genes related to *Trypanosoma cruzi* infection in the human host.** 2011. 125 p. (Ph. D. Thesis Biology of the interactions between Hosts and Pathogens). Biomedical Science Institute of the University of Sao Paulo, Sao Paulo, 2011.

Chagas disease is caused by the protozoan *Trypanosoma cruzi* and is usually transmitted by Triatominae bugs and affects about 12 million people in the American continent. In a previous study, segregation analysis showed evidence of a major gene with a small multifactorial component influencing the predisposition to the *Trypanosoma. cruzi* infection in a population composed by 4697 individuals of 886 families from Northeastern Brazil in 1969-1970 at São Paulo, Brazil. In the present work, 69 individuals (18 positives to *T. cruzi* infection and 51 negative) belonging to 14 families were selected. They had the DNA extracted and genotyped using 250K SNPs DNA microarrays (GeneChip Mapping Affymetrix). 18 SNPs showed evidence of association between infection to *T. cruzi* with $P < 10^{-5}$, although after Bonferroni's correction only the SNP rs17469997 (minor allele frequency = 0.1667, adjusted-Bonferroni $P = 0.015$) on chromosome 10 was significant. The other 17 SNPs that showed association with *T. cruzi* infection with $P < 10^{-5}$ can still be informative in linkage analyses. On an effort to validate these findings, a multi point linkage analyses was performed with GeneHunter (KRUGLYAK et al., 1996) program and a two point linkage analyses were performed with SuperLink (FISHELSON e GEIGER, 2002) program, both analyses showed no significant results.

Keywords: Chagas disease. genetic epidemiology. genechip.

SUMÁRIO

1 INTRODUÇÃO	12
2 ETIOLOGIA	20
2.1 Histórico	20
2.2 A Doença	23
2.3 O Ciclo	26
2.4 Área endêmica	28
2.5 Epidemiologia	30
3 OBJETIVOS	31
4 MATERIAIS E MÉTODOS	32
4.1 Casuística	32
4.1.1 Aspectos éticos da pesquisa	32
4.1.2 População	33
4.1.3 Diagnóstico sorológico	37
4.1.4 Reação de imunofluorescência indireta	38
4.1.5 Controle de qualidade da reação	39
4.2 Extração de DNA	40
4.3 Amostra estudada	43
4.4 Genotipagem com microarranjos de DNA	44
4.5 Dados	47
5 RESULTADOS	49
6 DISCUSSÃO	65
7 CONCLUSÕES	74
REFERÊNCIAS	75
APÊNDICE	83

1 INTRODUÇÃO

O estudo conjunto do genoma do hospedeiro e características sorológicas podem indicar a resistência/suscetibilidade à infecção por algumas doenças complexas como a doença de Chagas e o conhecimento sobre mecanismos de predisposição genética pode contribuir para o melhor entendimento, a prevenção e o tratamento dessas patologias. Baseando-se em fatos como: ocorrência de indivíduos com resistência ao contágio, associação familiar, variação racial da prevalência ou da incidência e associação entre polimorfismos e a moléstia, pode-se supor que certos mecanismos genéticos podem proporcionar resistência à doenças infecciosas (BEIGUELMAN, 1994).

Sabe-se que a capacidade de resistência/suscetibilidade humana a parasitos varia amplamente entre indivíduos e pode influenciar a progressão da infecção. Pesquisas na área de epidemiologia genética direcionam seus focos às doenças complexas que afetam populações devido a fatores genéticos e ambientais (BRACKEN, 2005).

A epidemiologia genética trata da etiologia, distribuição e controle de doenças em grupos de indivíduos aparentados e procura conhecer as causas herdáveis de doenças nas populações (MORTON, 1982). Dessa forma o núcleo do estudo de epidemiologia genética é a família, enquanto que conclusões válidas requerem estudos de populações maiores (RAMAGOPALAN; DYMENT; EBERS, 2008).

Vários trabalhos têm indicado a presença de mecanismos genéticos importantes agindo no fenótipo resistente/suscetível a doenças complexas como a tuberculose (KALLMANN; REISNER, 1942; COMSTOCK, 1978; BELLAMY et al., 2000; TOSH et al., 2006), hanseníase (BEIGUELMAN, 1965, 1968; CHAKRAVARTI; VOGEL, 1973; FEITOSA et al., 1995; FITNESS et al., 2004), hepatite B (LIN et al., 1989; HILL, 1996; JIANG; WANG; LIU; LIU, 2003), malária (JAMES; NICOL; SHUTE, 1932; ALLISON, 1954; FLORI et al., 2003; entre outros) e até HIV (CARRINGTON et al., 1999; LIU et al., 2006).

Outros estudos também têm mostrado uma grande base poligênica para a suscetibilidade para muitas doenças infecciosas comuns (KRIEGER; FEITOSA, 1999, HILL, 2001), embora dados da literatura também tenha indicado a evidência de um mecanismo genético simples para explicar a resistência e suscetibilidade a doenças infecciosas, entre elas a doença de Chagas (KRIEGER; FEITOSA, 1999; WILLIAMS-BLANGERO et al., 2003).

Indícios de mecanismos genéticos que atuam no fenótipo resistente/suscetível foram encontrados graças à descoberta e a utilização de marcadores moleculares para várias doenças complexas como a asma, câncer, diabetes e até para determinados aspectos de características cardiovasculares. Essas descobertas mostram que novas ferramentas da genética estão surgindo e são cada vez mais poderosas no tocante ao entendimento e tratamento de vários tipos de doenças.

Novas informações da sequência do genoma humano podem alterar muitos aspectos da prática clínica, aumentando nosso conhecimento dos mecanismos das

doenças e nos ajudando a identificar novos marcadores genéticos. A fenotipagem, que utiliza informações do genoma pode permitir predições e diagnósticos mais precisos e vem chamando a atenção de pesquisadores e gerando controvérsias sobre a maneira que essas metodologias irão afetar a prática da medicina do dia-a-dia (BELL, 2004).

Atualmente muitos estudos têm sido realizados com a utilização de SNPs (*single nucleotide polymorphism*) ou polimorfismos de base única que ocorrem quando as variações em um único ponto da sequência de DNA com frequência de pelo menos 1% do variante menos freqüente na população.

A aplicação de métodos de microarranjos no estudo de doenças complexas está se tornando cada vez mais barata e as plataformas de genotipagem cada vez mais sofisticadas.

A obtenção de DNA genômico para análises que utilizam microarranjos ou *genechips* envolve investimento, estrutura adequada, tempo e habilidade para a extração (JASMINE et al., 2008).

Com a técnica de microarranjos de DNA pode-se criar mapas de SNPs que podem ajudar a identificar genes associados às doenças. A análise desses microarranjos permite o estudo de centenas de milhares de polimorfismos de base única (SNPs) e vem sendo bastante utilizada em estudos de associação ampla do genoma (Genome Wide Association, GWA) para relacionar a variabilidade genética com a diversidade fenotípica. Os resultados desses estudos auxiliaram no entendimento de diversas doenças (FRAZER et al., 2009) e têm sido usados para

identificar regiões de interesse para várias características e doenças complexas através da análise de milhares de SNPs distribuídos pelo genoma humano (GOLDSTEIN et al., 2010; THE WELLCOME TRUST CASE CONTROL CONSORTIUM, 2007).

Entre os estudos que utilizam a técnica de microarranjos para estudar SNPs, trabalhos mais recentes ganharam destaque aplicando métodos de GWAS e identificando genes que podem ser responsáveis por doenças como a diabetes tipo 1 e doença de Chron (LEHNE et al., 2011), enxaqueca (LIGHTART, 2011) e outras doenças comuns (LIU et al., 2011).

Os primeiros estudos desse tipo encontraram associação do fator de complemento H com a degeneração macular (KLEIN et al, 2005), essa descoberta abriu as portas para outros estudos de varredura genômica sendo considerado como o início dos grandes estudos do tipo GWA. Seguindo essa metodologia, no ano de 2007 houve uma explosão no número de estudos do tipo GWA para várias doenças complexas (KU et al, 2010).

Um estudo realizado com duas populações, vindas da Colômbia e do Peru, mostrou que existe uma diferença significativa na distribuição entre indivíduos caso e controle do alelo 173C do gene MIF (fator inibitório da migração dos macrófagos) conferindo a suscetibilidade à doença de Chagas (TORRES, 2009).

O interferon gama (*IFNG*) também mostrou uma diferença significativa entre indivíduos caso e controle do alelo *IFNG* +874/A ($P=0,003$) (TORRES, 2010); o mesmo ocorreu com a distribuição dos alelos 10T e 10C do gene *TGF β 1* em um

teste do tipo caso-controle (CALZADA, 2009) nos estudos realizados para verificar a suscetibilidade à doença de Chagas.

Devido ao fato de que apenas alguns pacientes com a doença de Chagas desenvolvem cardiopatias, alguns pesquisadores focaram seus estudos para testes de associação, entre genes e a forma cardiopática da doença de Chagas, utilizando SNPs (RAMIREZ et al., 2010).

Apesar do presente estudo abordar apenas a infecção por *T. cruzi*, entre os genes associados com cardiopatias da doença de Chagas a homozigosidade dos alelos *LTA* +80A e *LTA* +252A é mais freqüente (RAMASAWMY, 2007); indivíduos com o gene *BAT1* juntamente com a reduzida expressão do HLA-B estão mais predispostos a desenvolver a cardiopatia da doença de Chagas (RAMASAWMY et al., 2006a). Ramasawmy et al em 2006, verificou que indivíduos com o genótipo *CCL2* _2518AA têm um risco quatro vezes maior de desenvolver a mesma cardiopatia (RAMASAWMY, 2006b) entre outros.

Da mesma forma, pacientes positivos para os alelos *TNF2* ou *TNFA2* mostraram menor taxa de sobrevivência a cardiopatia da doença de Chagas crônica que os controles apropriados (DRIGO, 2006).

Alguns SNPs da família TLR influenciam a resposta inflamatória e estão associados à suscetibilidade a infecções (SHRÖDER; SCHUMANN, 2005). Estudos mostram que existe uma associação entre a heterozigosidade nos SNPs da família TLR (Toll-like receptors) e o baixo risco de desenvolvimento de cardiopatias em indivíduos infectados por *T. cruzi* (RAMASAWMY et al., 2009).

Trabalhos anteriores de nosso laboratório evidenciaram a existência de fatores genéticos no controle dos fenótipos resistente/suscetível a infecção pelo *Trypanosoma cruzi*, através da utilização de métodos estatísticos em análise de famílias vindas de locais onde a doença está presente (FEITOSA, 1985; KAWAMATA, 2006).

Em um desses trabalhos foram estudados 4.697 indivíduos de 886 famílias provenientes principalmente das regiões Norte e Nordeste do Brasil que passaram pela extinta Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo, local onde hoje fica o Memorial do Imigrante.

Entrando em maiores particularidades desse trabalho, a prevalência da infecção por *T. cruzi* nessa população é de 5,79%. Parte dessa amostra foi estudada, sendo utilizadas informações sobre sexo, grau de parentesco, tipo de habitação, idade e diagnóstico sorológico. Essa amostra será detalhada mais adiante em Casuística (KAWAMATA et al., 2007).

Com o uso de programas auxiliares desenvolvidos no laboratório, do software SPSS e do programa POINTER (LALOUEL; MORTON, 1981) foram feitas as análises genético-estatísticas. Os resultados de análise de regressão múltipla escalonada da reação de imunofluorescência com as variáveis: idade, idade², sexo, sexo x idade, sexo x idade² e tipo de habitação indicaram que o melhor modelo apresentou coeficiente de determinação igual a $R^2 = 0,035$ e indicam que fatores como idade, idade² e tipo de habitação são responsáveis diretamente pela variabilidade da infecção por *T. cruzi* em cerca de 4%, enquanto outras variáveis como o sexo, sexo x idade e sexo x idade², podem ser desconsideradas (Anexo 1).

É importante relatar os resultados anteriormente obtidos na amostra aqui estudada (KAWAMATA, 2006) a fim de evidenciar a necessidade de aprofundamento desses estudos. A análise de correlação da infecção por *T. cruzi* nessa população mostrou valores significativos para todos os tipos de pares testados, Pai x Mãe, Pai x cada Filho, Pai x média dos Filhos, Mãe x cada Filho, Mãe x média dos Filhos, média dos Pais x cada Filho, média dos Pais x média dos Filhos e entre Filhos, sugerindo uma herdabilidade de 20% ou $h^2 = 0,2$ (Anexo 2). Como o par testado Pai x Mãe também se mostrou significativo, temos também a confirmação da influência do ambiente, que foi confirmada pelas análises de componentes de variância de Fisher (1918) indicando que essa influência gira em torno de 52,3% ou 0,523 (Anexo 3).

Foi realizada também, uma análise de segregação complexa (Anexo 4) que é mais apropriada que as análises de correlação por que analisa todos os elementos como um todo e não apenas em pares como a análise de correlação. Esta análise, revelou que a infecção por *T. cruzi* depende, além de um componente multifatorial significativo, de um gene dominante, indicando a existência de um mecanismo genético principal importante associado a essa infecção. Estes resultados sugerem a existência de um gene com frequência de $q=0,605$, responsável por cerca de 12% da variabilidade desta característica, fato que corrobora outros estudos que sugerem a presença de um mecanismo genético relacionado com a infecção por *T. cruzi* (KAWAMATA, 2006).

Devido a estes fatos, o presente trabalho visa à busca de marcadores genéticos que possam indicar genes responsáveis pela resistência/suscetibilidade a

infecção por *Trypanosoma cruzi* nos indivíduos que passaram pela Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo, utilizando a metodologia de epidemiologia genética aplicada a técnica de microarranjos de DNA.

Amostras de sangue e saliva de todos os indivíduos pesquisados no trabalho acima (KAWAMATA, 2006) se encontram estocadas em congeladores com temperatura a cerca de -80 °C, no Laboratório de Epidemiologia Genética, do Departamento de Parasitologia da Universidade de São Paulo (USP), juntamente com o cadastro dos indivíduos, o que favorece o prosseguimento desses estudos.

2 ETIOLOGIA

2.1 Histórico

A Doença de Chagas foi descoberta em 1909 por Carlos Chagas (1878-1934), pesquisador do Instituto de Manguinhos. Por designação de Oswaldo Cruz, foi responsável pela equipe que combateria uma epidemia de malária que atingia trabalhadores da Estrada de Ferro Central do Brasil em Lassance, Minas Gerais.

Carlos Chagas trabalhando em seu vagão-laboratório, que também utilizava como dormitório, buscava novos parasitos e artrópodes que poderiam servir como vetores. Foi então que o chefe da comissão dos engenheiros mostrou-lhe um percevejo hematófago muito comum que vivia nas frestas das casas de pau-a-pique da região. Apelidado de barbeiro por picar preferencialmente o rosto das pessoas enquanto dormiam.

Ao examinar o barbeiro, Carlos Chagas encontrou um protozoário flagelado que pensou ser o *Trypanosoma minasense*, que ele mesmo havia identificado como uma nova espécie não-patogênica presente no sangue de sagüis, sendo então o barbeiro, o inseto vetor. Após alguns testes foi constatado que não se tratava do *Trypanosoma minasense* e sim de uma nova espécie que Carlos Chagas batizou de *Trypanosoma cruzi* em homenagem ao seu mentor Oswaldo Cruz.

De volta ao Instituto de Manguinhos, Carlos Chagas iniciou estudos sobre o ciclo evolutivo do novo parasita e verificou que o mesmo era capaz de infectar

animais vertebrados. Suspeitando que o ser humano pudesse ser um desses vertebrados que potencialmente poderia ser infectado, Chagas retornou a Lassance para realizar exames de sangue nos moradores.

Durante esses exames encontrou o parasita *Trypanosoma cruzi* no sangue de uma criança febril: era Berenice, de dois anos de idade que se tornou o primeiro caso conhecido da nova doença humana. Berenice faleceu em 1981 devido a insuficiência cardíaca aos 73 anos de idade, sendo que durante 70 anos conviveu com a doença de Chagas (FIOCRUZ, 2011).

Fato único na história, Carlos Chagas foi responsável pelo estabelecimento da etiologia, ciclo parasitário, identificação do inseto vetor, identificação dos reservatórios domésticos e silvestres, descrição da doença e seu diagnóstico, construindo assim todas as facetas da nova doença (Rey, 2001).

Esta descoberta deu grande impulso a carreira científica de Carlos Chagas. Em março de 1910 ele assumiu o cargo de “chefe de serviço” que era considerado como cargo de sucessão de Oswaldo Cruz. Em 26 de outubro de 1910, se tornou membro titular da Academia Nacional de Medicina.

No ano seguinte, um evento marcou a divulgação da descoberta e da nova doença no cenário científico internacional, projetando o Instituto Oswaldo Cruz como referência em pesquisas de doenças tropicais principalmente a Doença de Chagas.

Em 1912, Carlos Chagas conquistou o Prêmio Schaudinn, concedido de quatro em quatro anos pelo Instituto de Doenças Tropicais de Hamburgo ao melhor trabalho em protozoologia.

Com a grande repercussão pela descoberta e estudos da Doença de Chagas, Oswaldo Cruz conseguiu verba do governo federal para equipar um hospital em Lassance,

Sob a liderança de Chagas e com a colaboração de vários pesquisadores do Instituto Oswaldo Cruz (como Gaspar Vianna, Arthur Neiva, Astrogildo Machado, Eurico Villela, Carlos Bastos de Magarinos Torres, entre outros), a nova tripanossomíase passou a ser estudada em seus mais diferentes aspectos, tais como as características biológicas do vetor, do parasito e de seu ciclo evolutivo, o quadro clínico e a patogenia, as características epidemiológicas, os mecanismos de transmissão e as técnicas de diagnóstico. (FIOCRUZ, 2011).

2.2 A Doença

Apesar da doença de Chagas não ser o tema central do presente trabalho e sim a infecção do agente etiológico principal, apresentamos alguns aspectos relevantes da doença.

A doença de Chagas é uma doença parasitária causada pelo protozoário flagelado *Trypanosoma cruzi*.

A tripanossomíase americana, como também pode ser chamada a doença de Chagas, ocorre em alguns lugares do Brasil e do mundo, é transmitida principalmente por insetos vetores da subfamília Triatominae (Hemiptera, Reduviidae). Conhecido popularmente como barbeiro ou chupança, esse inseto se aninha nos domicílios, principalmente em fendas na parede e teto de casas de pau-a-pique e residências precárias da área rural e de favelas da periferia, podendo também ser encontrado atrás de mobílias. O habitat do barbeiro é muito favorecido nesses tipos de construções de baixa qualidade feitas de madeira e barro, relacionando diretamente a doença de Chagas à pobreza.

A doença atinge humanos e um grande número de espécies de animais domésticos e selvagens. Os barbeiros adquirem o *Trypanosoma cruzi* ao ingerir sangue de animais ou humanos infectados; ao sugar (muito raramente) o conteúdo intestinal de outro barbeiro infectado, ou ao se alimentar de fezes infectadas de outro barbeiro (coprofagia). Não há transmissão do *Trypanosoma cruzi* de fêmeas de barbeiro para a prole: os ovos não dão origem a ninfas infectadas.

A forma clássica de infecção se dá por contato com as fezes do inseto vetor infectado no momento da picada, onde eventualmente o hospedeiro tem a tendência de coçar o local da picada inoculando, assim, o *Trypanosoma cruzi* na sua corrente sanguínea (REY, 2001).

Existem outras formas de infecção que podem ocorrer através de alimentos contaminados, transfusões de sangue, transplante de órgãos, transmissão congênita e até acidentes de laboratório.

Os estudiosos separam a infecção chagásica em duas etapas: a fase aguda e a fase crônica. A fase aguda pode ser assintomática ou apresentar sintomas como febre, astenia, poliadenite, aumento do fígado, aumento do baço, sinal de Romaña e outros tipos de chagomas de inoculação. Nesse período os tripanossomos são facilmente encontrados na corrente sangüínea. Desses, dez por cento chegam a falecer de morte súbita, meningocefalite e, mais raramente, falência cardíaca devido à miocardite aguda difusa. As alterações cardíacas da fase aguda costumam ser reversíveis, mas em crianças com menos de cinco anos a mortalidade é alta (Rey, 2001).

Após a fase aguda vem uma fase indeterminada, onde o paciente pode ficar longos períodos que podem ultrapassar 30 anos sem apresentar sintoma algum. Essa fase afeta de 50 a 70% dos indivíduos infectados, sendo que a infecção só é perceptível nessa fase através de exames clínicos detalhados, reações sorológicas ou outros métodos laboratoriais.

Durante a fase crônica alguns pacientes começam a apresentar sintomas como insuficiência cardíaca congestiva, falta de coordenação motora, megaesôfago, megacólon, alterações psicológicas, comportamentais e falta de memória, mesmo com a baixa frequência do parasito no hospedeiro nessa fase. Os danos aos órgãos são irreversíveis e alguns pacientes chegam a ter morte súbita. Em alguns casos a fase crônica sintomática pode vir imediatamente depois da fase aguda, sem a fase indeterminada assintomática e às vezes até mesmo sem a fase aguda (Rey, 2001).

Para o tratamento dos pacientes, somente o benznidazol e o nifurtimox são indicados dentre as numerosas drogas já testadas. O efeito dessas drogas varia de acordo com o próprio paciente e da linhagem do parasito e em alguns casos, pacientes tiveram o diagnóstico negativedo. Devido a efeitos colaterais importantes é necessário o consentimento esclarecido e acompanhamento laboratorial do paciente.

Outras terapias visam os sintomas. Para a cardiopatia utilizam-se medicamentos como o amiodarone, a propafenona e a mexiletine. Para os casos de megaesôfago e megacólon o tratamento é cirurgico.

2.3 O Ciclo

O ciclo do parasito (Figura 1) no barbeiro começa com o inseto ao picar um hospedeiro vertebrado infectado adquirindo, assim, a forma tripomastigota do protozoário. Na parte posterior do intestino do inseto a forma tripomastigota se transforma em epimastigota, estágio em que o protozoário se multiplica. A seguir, na parte do reto do intestino do inseto, a forma epimastigota passa a ser tripomastigota metacíclica que é a forma infectante do protozoário.

No hospedeiro vertebrado, o ciclo tem início quando o inseto o pica em busca de seu sangue, neste momento o barbeiro elimina fezes contaminadas juntamente com o protozoário na forma tripomastigota metacíclico ao lado da picada. Com o auxílio do hospedeiro, ou não, ocorre a penetração das fezes juntamente com o *Trypanosoma cruzi* na sua corrente sanguínea. A forma tripomastigota metacíclico ao invadir as células sanguíneas passa para a forma amastigota.

Uma vez dentro da célula, a forma amastigota se multiplica inúmeras vezes até romper a parede da célula e cair novamente na corrente sanguínea retornando para a forma tripomastigota metacíclico. Na corrente sanguínea o protozoário está livre para invadir outras células do hospedeiro ou ser sugado para dentro de outro inseto vetor, completando assim, o ciclo.

Figura 1 – Ciclo da doença de Chagas

Trypanosomiasis, American (Chagas disease) (*Trypanosoma cruzi*)

Fonte: Center for Disease Control and Prevention

2.4 Área endêmica

A área endêmica da doença de Chagas se estende por toda a América Latina, desde a fronteira do sul dos Estados Unidos com o norte do México até o sul da Argentina (figura 2).

Casos fora da área endêmica foram registrados em países da Europa como a Espanha e países da América no Norte como Canadá e Estados Unidos (FIOCRUZ, 2011). Acredita-se que esses casos esporádicos se devem a transplantes de órgãos e de sangue doados por indivíduos que passaram pelas áreas endêmicas e se tornaram portadores da infecção pelo *Trypanosoma cruzi*.

Os movimentos de migração rural/urbana que ocorreram na América Latina nas décadas de 70 e 80 mudaram o padrão epidemiológico tradicional da doença de Chagas e a tornou uma infecção urbana rara, cuja forma mais importante de transmissão é a transfusional. As quantidades de sangue infectado nos bancos de sangue em algumas cidades selecionadas variam de 3 a 53%, mostrando que a prevalência da infecção por *T. cruzi* é maior do que a infecção por HIV e por hepatite B e C (FIOCRUZ, 2011).

Atualmente, há uma tendência ao decréscimo contínuo na prevalência da infestação doméstica pelo inseto vetor (*Triatoma infestans*) e na incidência de infecção humana em crianças e jovens nos países do cone sul (FEITOSA e KRIEGER, 1991; ARAS et al., 2003; WHO, 2011).

Figura 2 – Mapa da área endêmica da doença de Chagas

Adaptado de OPAS, 2006

2.5 Epidemiologia

A doença de Chagas atinge aproximadamente de 10 a 14 milhões de pessoas, 25 milhões de pessoas vivem nas áreas de risco e estima-se que em 2008 mais de 10 milhões de pessoas morreram devido às complicações pela infecção pelo *Trypanosoma cruzi*. (FIOCRUZ, 2011; WHO, 2011).

É a quarta maior causa de mortes no Brasil superando as mortes decorrentes da infecção pelo vírus HIV e ficando atrás apenas da hepatite B, hepatite C e sífilis (FIOCRUZ, 2011).

Para tentar combater e ampliar o reconhecimento da importância dessa infecção, uma Rede Global para Eliminação da doença de Chagas (Global Network for Chagas Elimination – GNChE) foi criada e iniciou suas atividades em 2008. Coordenado pelo Instituto Oswaldo Cruz (IOC) esse programa tem o objetivo de combater e alertar sobre os riscos da doença de Chagas e estender a conscientização para todo o mundo, incluindo os países não-endêmicos.

3 OBJETIVOS

Tendo em vista os resultados de trabalhos anteriores a respeito da resistência/suscetibilidade a infecção por *Trypanosoma cruzi*, na amostra proveniente da Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo, essa pesquisa tem dois objetivos.

a) O objetivo inicial é recuperar e verificar a viabilidade de uso da amostra da Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo, datada de 1969-1970 para a realização da genotipagem utilizando a plataforma de microarranjos GeneChip.

b) Em sendo viável, tentar localizar possíveis genes que atuam nessa infecção no ser humano, tanto já observadas em trabalhos anteriores, como outros inéditos, através da utilização de métodos da genética epidemiológica.

4 MATERIAIS E MÉTODOS

4.1 Casuística

4.1.1 Aspectos éticos da pesquisa

Este estudo seguiu todas as normas de conduta e documentação da Comissão de Ética em Pesquisa do Instituto de Ciências Biomédicas da Universidade de São Paulo segundo o parecer 687/CEP.

Este trabalho está inserido no projeto “Dinâmica Gênica em Populações do Nordeste brasileiro” sob a responsabilidade do Professor Doutor Henrique Krieger.

Hoje as amostras coletadas neste projeto estão sob os cuidados do Laboratório de Epidemiologia Genética do Instituto de Ciências Biomédicas da Universidade de São Paulo, devidamente estocadas em freezer -80 °C.

4.1.2 População

A amostra é proveniente de indivíduos que passaram pela Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo, situada no bairro da Mooca onde hoje fica o Memorial do Imigrante (Figura 3). Era um local destinado a receber imigrantes de outros países ou de outros estados, a fim de estabelecê-los nas lavouras, nos núcleos coloniais ou na indústria. Com inauguração no ano de 1908 a Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo recebeu aproximadamente 3 milhões de pessoas durante 91 anos de funcionamento.

Uma equipe dirigida pelo Prof. Dr. Henrique Krieger (projeto: Dinâmica Gênica em Populações do Nordeste Brasileiro) coletou amostras de 9.127 indivíduos de 1806 famílias vindas predominantemente do nordeste brasileiro no período de outubro de 1.969 a agosto de 1.970.

Essa população caracteriza-se por apresentar altas taxas de fertilidade, consangüinidade, mortalidade e grande variabilidade étnica permitindo a realização de estudos de variabilidade genética. Os critérios de inclusão foram: (i) presença de ambos os cônjuges (famílias nucleares), (ii) origem rural (possibilitando a homogeneidade da amostra) e (iii) idade materna inferior a 50 anos (para maior confiança nos dados relativos a mortalidade fetal).

Figura 3 – Memorial do Imigrante, extinta Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo

Fonte: Wikimedia Communs - Lukaaz

Os indivíduos foram numerados e passaram por entrevista, onde foram obtidas informações como: origem, parentesco, tempo de coabitação, local e data de nascimento, distância entre os locais de nascimento, história reprodutiva e nível social. Sendo que o nível social foi auferido por alfabetização e tipo de habitação (tipo de material da edificação e condições sanitárias) O tipo de habitação foi dividido em três grupos; (i) casa de barro, sopapo, taipa, tábua; (ii) casa de adobe rebocado ou tijolo, sem privada ou chão; (iii) casa rebocada com privada e chão. A classificação étnica também foi feita, levando-se em conta a cor da pele, a conformação dos lábios, do nariz e o tipo de cabelo.

Após a entrevista, os indivíduos foram submetidos a um exame médico, onde passaram por uma anamnese e exame físico com a verificação da pele e tecido subcutâneo, ausculta cardíaca, inspeção e palpação do abdome e exame do sistema locomotor. Foram coletados 10 ml de sangue e cerca de 5 ml de saliva e foram medidas a altura, o peso, as pressões arteriais e o pulso radial.

O coeficiente de endocruzamento dessa população, através de métodos de bioensaio (uso de marcadores genéticos), foi estimado por Cabello (1972), obtendo para a geração paterna $F = 0,0227 \pm 0,00376$ e para a geração filial $F = 0,0262 \pm 0,00327$.

Cabello (1976), também, apresentou alguns dados que caracterizam demograficamente esta população, tais como: número médio de gestações por casal igual a 5,32; a mortalidade pré-natal atingindo 10,93% e a mortalidade infantil 17,23%; e finalmente a alta taxa de analfabetismo (68%). A mistura étnica foi medida através da proporção de genes negróides, caucasóides e indígenas, fazendo uso de diversos polimorfismos genéticos (KRIEGER et al., 1965; RUSSO, 1972). Essa última autora confirmou o alto grau de mistura étnica, sendo composto por $0,354 \pm 0,014$ de genes negróides, $0,095 \pm 0,043$ de genes de ameríndios e $0,551 \pm 0,117$ de genes caucasóides, além de concluir que esta população já percorreu 95,7% do caminho em direção à panmixia étnica. Convém salientar que Penalva da Silva (1971), baseado em dermatóglifos chegou a estimativas de mistura étnica semelhantes.

Dal Colletto (1972) estudou a taxa de ilegitimidade paterna, materna e taxa de adoção dessas famílias, baseando-se nas frequências genotípicas para os

sistemas genéticos ABO, MN e Hp e pelas informações prestadas pelos cônjuges, através do método estatístico de verossimilhança máxima, obtendo uma baixa taxa de ilegitimidade.

Outras informações de interesse médico, genético e antropológico podem ser encontradas em Barbosa (1971, 1976), Barbosa, Koury e Krieger (1975), Penalva da Silva (1975), Mestriner (1976), Moro-Furlani (1977), Dal Colletto, Krieger, Magalhães (1981) e Penalva da Siva e Krieger (1981, 1982).

Os testes de reação de imunofluorescência indireta para a determinação da presença do parasita *Trypanosoma cruzi* no hospedeiro foram realizados pela Dra. Mary Furlan Feitosa (1985).

4.1.3 Diagnóstico sorológico

O diagnóstico sorológico foi efetuado pela Dra. Mary Furlan Feitosa, segundo a técnica desenvolvida por Camargo (1966, 1975), utilizando a cultura do *Trypanosoma cruzi* na forma epimastigota em meio LIT (Liver Infusion Tryptose) (CAMARGO, 1964).

4.1.4 Reação de imunofluorescência indireta

Na técnica de imunofluorescência indireta incuba-se o antígeno com o anticorpo específico (imunoglobulina) formando uma estrutura antígeno-anticorpo. Depois da lavagem e é adicionada uma anti-imunoglobulina fluorescente que se liga a essa estrutura antígeno-anticorpo formando uma estrutura antígeno-anticorpo-anti-imunoglobulina fluorescente. Uma nova lavagem é realizada e a seguir é realizada a observação em microscópio de fluorescência. Essa técnica é muito utilizada pois amplifica o sinal luminoso por aumentar o número de sítios através do anticorpo específico, aumentando a sensibilidade e ampliando a fluorescência (cf. REY, 2001).

4.1.5 Controle de qualidade da reação

Inicialmente as lâminas de fluorescência para os testes laboratoriais de controle de qualidade de reação para a doença de chagas foram preparadas no Departamento de Parasitologia da Escola Paulista de Medicina (EPM) pela Dra. Mary Furlan Feitosa conforme descrito abaixo. E um segundo momento, as mesmas foram posteriormente substituídas pelas lâminas industrializadas do Centro de Controle e Produtos para Diagnósticos (CECON) permitindo maior eficiência.

Com o objetivo de verificar a confiabilidade do método, foram testados cegamente 246 indivíduos simultaneamente, com as duas fontes de antígenos. Desses indivíduos 17 apresentaram reação positiva e 229 mostraram reação negativa, para ambas as lâminas, não se evidenciando nenhuma discrepância. Dessa forma, concluiu-se pela completa equivalência entre as duas fontes fornecedoras das lâminas-testes permitindo portanto, a soma dos resultados dos dois tipos de testes e fundamentalmente uma constatação de alta reprodutibilidade dos resultados. Uma análise para comparar as proporções de respostas dos indivíduos examinados com o material fornecido pela EPM e aqueles que foram testados com o material da CECON reforçou essa conclusão (Tabela 1).

Tabela 1 – Comparação entre as duas fontes de lâminas de imunofluorescência

	EPM	CECON	TOTAL
NEGATIVO	430	4660	5090
POSITIVO	32 (6,9%)	299 (6,0%)	331 (6,1%)
TOTAL	462	4959	5421

$$\chi^2 = 0,595 \text{ G.L.} = 1; 0,02 < P < 0,01$$

4.2 Extração de DNA

As amostras de soro e solução de hemácias glicerolizadas estão conservadas em tubos tipo eppendorf e estocados em freezer -80°C no Laboratório de Epidemiologia Genética do Departamento de Parasitologia no Instituto de Ciências Biomédicas da Universidade de São Paulo.

A extração do DNA de cerca de 400 amostras foi realizada a partir de células nucleadas encontradas em uma solução de hemácias, de acordo com o protocolo a seguir:

- Aliquotar 500 µl de sangue e adicionar 400 µl de Lise I e 200 µl de fenol.
- Homogeneizar levemente e centrifugar por 15 minutos a 12.000 g a temperatura ambiente.
- Recuperar o sobrenadante.
- Adicionar 400 µl de clorofórmio: álcool isoamílico (24:1).
- Centrifugar por 15 minutos a 12.000 g a temperatura ambiente.
- Adicione 0,1 volumes de acetato de sódio 3M + 3 volumes de etanol 100% + 1000 µl glicogênio.
- Incubar por 16 horas a -20 °C.
- Centrifugue por 15 minutos a 12.000 g a temperatura ambiente.
- Descarte o sobrenadante.
- Lave com 1000 µl de etanol 70% duas vezes ou mais centrifugando a 12.000 g a temperatura ambiente.

- Secar totalmente na capela.
- Ressuspender com 50 µl de H₂O Miliq ou TE.

Preparação da solução de Lise I:

- 1,21 g de Tris/HCl 0,01 M pH 7,6
- 109,54 g de sacarose 0,32 M
- 0,5 ml de cloreto de magnésio 5 mM (5 ml MgCl₂ 1 M) (0,5 ml)
- 1 ml de Triton X-100 1%
- Completar o volume para 100 ml com água destilada

Preparação de TE:

- 10 ml de 1M Tris
- 2 ml de EDTA 0,5 M
- Completar o volume para 1 l com água destilada

A seguir, a quantificação de 1 µl de amostra foi realizada com o equipamento de espectrofotometria Thermo Scientific NanoDrop™ 2000 Spectrophotometer.

Após essa quantificação do DNA, verificou-se que a quantidade de DNA nas amostras não seria suficiente para os estudos com o GeneChip, pois não atingiu a concentração de 50 ng/µl. Utilizou-se então a enzima Phi 29 (Illustra GenomiPhi™

V2 DNA Amplification Kit), com o intuito de se realizar um WGA (*whole genome amplification*) o ou amplificação total do genoma de forma a se obter a quantidade necessária de DNA para a realização do estudo.

A amplificação foi feita de acordo com o protocolo a seguir.

- Utilizar 1 µl (1-10 ng) de DNA a ser amplificado.
- Adicionar 9 µl do tampão de amostra do GenomiPhi DNA Amplification Kit.
- Denaturar a 95 °C por 3 minutos e resfriar imediatamente a 4 °C.
- Adicionar 9 µl do tampão de reação do GenomiPhi DNA Amplification Kit e 1 µl da enzima PHI 29.
- Incubar a 30 °C por 1,5 horas e a seguir aquecer a 65 °C por 10 minutos para inativar a enzima.

4.3 Amostra estudada

Neste trabalho foram selecionados amostras de sangue de 69 indivíduos vindos da Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo. Eles foram selecionados por serem pertencentes a famílias nucleares completas com pelo menos um dos pais infectado por *Trypanossoma cruzi*.

Estes indivíduos têm faixa etária variando de 1 até 40 anos de idade, são pertencentes a 14 famílias sendo que 18 são positivos para a infecção pelo *Trypanosoma cruzi* e 51 são positivos para essa infecção, independentemente do grau da infecção ou do estágio da doença de Chagas.

4.4 Genotipagem com microarranjos de DNA

A realização da genotipagem com microarranjos de DNA seguiu os padrões recomendados pela empresa fornecedora da plataforma (Affymetrix). Para a manipulação dos reagentes e dos microarranjos, é necessário ter um laboratório bem estruturado com locais adequados para cada fase do protocolo. Para isso, o trabalho de bancada da genotipagem foi realizado em dois locais separados, o laboratório de uso comum e uma sala limpa separada.

No laboratório ficam os equipamentos de rotina diária comum a todo laboratório como bancada, pipeta, termocilador, centrifuga, banho seco, fluxo laminar, freezer e os equipamentos da plataforma de genotipagem que são: forno de hibridação, estação de lavagem, *scanner* e computador.

Com o objetivo de evitar contaminações, a sala limpa é composta por uma pré-sala, onde o pesquisador deve vestir touca de cabelo, avental, protetor dos pés, mascara e luvas, todos estéreis e descartáveis. A segunda sala é composta por bancada, pipeta, fluxo laminar, centrífuga e termociclador. A última sala é o local onde o pesquisador retira as vestimentas descartáveis. A estrutura da sala limpa é de fluxo único, ou seja, uma porta de entrada e outra de saída, sendo que o pesquisador não deve retornar de uma sala para outra para evitar a contaminação do local.

Com a estrutura pronta, foi dada seqüência ao protocolo de genotipagem (Anexo 5) que tem início com a normalização das amostra para concentração de 50 ng/μl com tampão TE.

A seguir foi feita a digestão das amostras em locais específicos com a enzima NSP I (NEB biolabs) e posterior ligação (T4 DNA Ligase) com adaptadores (GeneChip Human Mapping 250K NSP Assay Kit) preparando a amostra para a reação de PCR (reação de polimerase em cadeia), onde o material será amplificado.

Neste ponto um gel de comum de rotina de agarose 2% deve ser feito para a averiguação da distribuição do produto da PCR. As bandas devem estar entre 250 bp e 1100 bp, confirmando o sucesso da digestão, ligação e PCR permitindo a seqüência do protocolo.

O produto da PCR foi purificado utilizando o DNA amplification Clean-Up Kit (Clontech). A seguir o material foi eluido em tampão (RB, Clontech), quantificado e normalizado para 2 μg/μl.

O material é fragmentado com o reagente de fragmentação (GeneChip Human Mapping 250K NSP Assay Kit) e um novo gel de agarose 4% foi feito com o intuito de verificar se o material está devidamente fragmentado com tamanho menor que 180 pares de base.

Após o sucesso da fragmentação o material foi marcado com GeneChip® DNA Labeling Reagent (30 mM) biotina, ficando adequado para a hibridação com o microarranjo GeneChip Human Mapping 250K NSP Array.

Depois de 16 horas de hibridação a 50 °C, o chip foi lavado para que o material não específico e em excesso seja removido. Nesse ponto, uma amplificação do sinal é realizada através de ciclos de incubações com estreptavidina-ficoeritrina (SAPE – Life Technologies – 1 mg/ml) e anticorpo anti-estreptavidina biotinilado (Vector – 0,5 mg/ml).

O último passo é a leitura do chip em um *scanner* (GeneChip Scanner 3000 7G), onde são utilizados programas específicos para a captação de imagem do GeneChip Human Mapping 250K NSP Array analisado.

A imagem dessa leitura pode ser observada com o programa de visualização Affymetrix Gene Chip Viewer, neste programa também pode-se verificar os controles positivos do sistema de forma a se comprovar que as condições de hibridação do microarranjo foram adequadas.

4.5 Dados

A genotipagem com o programa Affymetrix Genotyping Console é realizada a partir de arquivos .CEL que possuem dados de intensidade dos pixels coletados pelo *scanner*. Este mapa de pixels é convertido em genótipos criando o arquivo .CHP. Um terceiro arquivo foi montado em planilha Microsoft Excel onde as colunas representam o pedigree das famílias contendo informações a respeito da identificação do indivíduo, número da família, gênero, identificação dos pais (no caso dos filhos), ordem na família (pai, mãe ou filho), e a informação se o indivíduo é afetado ou não é afetado pela infecção do *Trypanosoma cruzi*.

O arquivo .CHP permite ao programa Affymetrix Genotyping Console converter os genótipos aferidos em outros dois arquivos com as extensões .ped e .map que contém todos os dados referentes à genotipagem do microarranjo de cada indivíduo no formato apropriado para a utilização no programa Plink (<http://pngu.mgh.harvard.edu/purcell/plink>).

Devido ao grande volume dos dados foi usado um segundo computador baseado no sistema operacional Linux (<http://ubuntu-br.org>) com maior capacidade de processamento onde foram realizados os testes de associação com o programa Plink e visualização dos resultados em gráficos e tabelas com o programa Haploview (<http://www.broad.mit.edu/mpg/haploview>).

Os dados também foram montados em um arquivo de texto (.txt) nos moldes para a utilização no programa GENIOC para os testes de mistura racial baseada

em marcadores clássicos (CABELLO e KRIEGER, 1997). Esse arquivo foi dividido em dois novos, um para os indivíduos do grupo caso e outro para os indivíduos do grupo controle.

Para os testes de ligação gênica foi utilizado o programa easyLINKAGE (HOFFMANN e LINDNER, 2005) que possui uma interface com opções de vários programas de ligação gênica, entre eles o GeneHunter (KRUGLYAK et al., 1996) e o SuperLink (FISHELSON e GEIGER, 2002). No entanto, para utilizar estes programas são necessários três novos arquivos.

O primeiro é um mapa dos SNPs presentes no GeneChip Human Mapping 250K NSP Array, esse arquivo é fornecido pela Affymetrix, empresa que produz os chips. O segundo arquivo, com o pedigree das famílias, foi montado no Microsoft Excel das famílias que compõe a amostra e contém a identificação do indivíduo, número da família, identificação dos pais (no caso dos filhos), gênero e se o indivíduo é afetado ou não é afetado pela infecção do *Trypanosoma cruzi* e salvo com a extensão .ped. E o terceiro arquivo também montado em Microsoft Excel, possui o genótipo de cada indivíduo para cada SNP presente no GeneChip de 260k e salvo com a extensão .snp.

Foram utilizadas as opções do programa GeneHunter (KRUGLYAK et al., 1996) para a ligação multi-ponto e o programa SuperLink (FISHELSON e GEIGER, 2002) para a ligação dois-pontos. Os resultados desses programas foram dados em arquivo tipo texto (.txt) e arquivo .pdf e .ps que podem ser visualizados com o programa Acrobat Reader e Acrobat Distiller (Adobe), respectivamente.

5 RESULTADOS

Os resultados da extração de DNA da amostra da Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo, após muitas tentativas e modificações no protocolo de extração (como descrito no item 4.2 – Extração de DNA), se mostraram bem sucedidos. Para se comprovar a existência de DNA, uma série de PCRs foi realizada com primers de STR especificamente humanos que foram visualizados em gel de eletroforese de agarose 4% (Figura 4 e Figura 5).

Figura 4 – Foto gel de verificação 1

Fonte: Laboratório de Epidemiologia Genética

Figura 5 – Foto gel de verificação 2

Fonte: Laboratório de Epidemiologia Genética

Porém, as amostras revelaram uma baixa concentração de DNA. Foi utilizada, então, a enzima PHI 29 (Ilustra GenomiPhi™ DNA Amplification Kit) para aumentar a concentração desse DNA.

Os valores da quantificação estão na Tabela a seguir.

Tabela 2 - Concentração do DNA de cada indivíduo

Indivíduo	Concentração de DNA		Indivíduo	Concentração de DNA
68	313,2 ng/μl		990	224,8 ng/μl
70	206,9 ng/μl		991	319,6 ng/μl
78	484,3 ng/μl		993	112,1 ng/μl
81	300,2 ng/μl		994	689,0 ng/μl
87	675,9 ng/μl		996	317,6 ng/μl
92	277,9 ng/μl		998	444,3 ng/μl
494	313,6 ng/μl		1000	466,6 ng/μl
500	158,2 ng/μl		1640	869,0 ng/μl
503	153,0 ng/μl		1651	217,8 ng/μl
505	420,3 ng/μl		1659	174,7 ng/μl
607	116,9 ng/μl		1749	340,6 ng/μl
611	252,9 ng/μl		1752	551,2 ng/μl
615	880,5 ng/μl		1755	404,8 ng/μl
620	276,6 ng/μl		1758	329,3 ng/μl
749	148,0 ng/μl		1761	204,1 ng/μl
766	281,1 ng/μl		1764	216,6 ng/μl
773	346,1 ng/μl		1767	207,1 ng/μl
782	430,0 ng/μl		6018	958,7 ng/μl
798	2519,4 ng/μl		6021	234,9 ng/μl
950	281,1 ng/μl		6025	212,7 ng/μl
953	316,2 ng/μl		6033	239,4 ng/μl
955	289,8 ng/μl		6035	269,9 ng/μl
959	767,1 ng/μl		6038	1130,2 ng/μl
971	409,5 ng/μl		6040	317,4 ng/μl
973	240,6 ng/μl		6045	332,1 ng/μl
974	722,0 ng/μl		6308	315,8 ng/μl
975	71,5 ng/μl		6312	395,0 ng/μl
977	200,2 ng/μl		6315	499,9 ng/μl
979	195,7 ng/μl		6320	186,7 ng/μl
981	242,0 ng/μl		6322	222,6 ng/μl
982	232,9 ng/μl		6328	215,4 ng/μl
983	252,8 ng/μl		6330	809,8 ng/μl
986	295,4 ng/μl		6333	143,5 ng/μl
987	468,8 ng/μl		6335	603,8 ng/μl
989	473,6 ng/μl		6038	113,2 ng/μl

Com as amostras em quantidade suficiente de DNA, foi dado seguimento ao protocolo de genotipagem com microarranjos.

Utilizando o programa PLINK (<http://pngu.mgh.harvard.edu/purcell/plink>), após a averiguação de legitimidade dos familiares, um estudo de associação foi realizado com 69 indivíduos genotipados da amostra da Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo. Destes, 18 indivíduos infectados pelo *Trypanosoma cruzi* e 51 indivíduos não infectados pelo *T. cruzi*, sendo 34 do sexo masculino e 35 do sexo feminino.

A Figura 6 abaixo mostra o *Manhattan Plot* com a representação gráfica dos resultados do teste de associação, comparando casos e controles para todos os 260 K SNPs.

Figura 6 - Gráfico do estudo de associação de 260 K SNPs

Nesse gráfico é possível observar que dos 260 K SNPs, os testes de associação apontaram para 18 SNPs que chamaram a atenção pelo baixo valor de P ($P < 10^{-5}$). Eles estão representados pelos pontos coloridos acima da linha azul do gráfico.

Os 18 SNPs mais significativos estão resumidos na Tabela 3 abaixo, de acordo com o valor de P crescente. Onde CRM = cromossomo, # SNP = número de referência do SNP, POSIÇÃO = referente à posição do SNP no cromossomo em questão, A1 = alelo menos freqüente, F_A = freqüência do alelo A1 nos indivíduos afetados, F_U = freqüência do alelo A1 nos indivíduos não afetados, A2 = alelo mais freqüente, X^2 = valor do qui-quadrado, P = probabilidade das diferenças entre caso e controle ser devida ao acaso e OR = *odd ratio*.

Tabela 3 – 18 SNPs mais significativos do teste de associação, por ordem crescente de P.

CRM	# SNP	POSIÇÃO	A1	F_A	F_U	A2	X^2	P	OR
10	rs17469997	8495938	C	0,3125	0,0000	G	29,36	0,00000006	-
7	rs1568799	42634401	A	0,2667	0,7895	C	25,56	0,00000043	0,0970
11	rs4923076	23203347	T	0,6000	0,1556	A	22,73	0,00000187	8,1430
X	rs4827928	97413650	A	0,8947	0,2836	G	22,72	0,00000188	21,4700
4	rs2086048	54221234	A	0,6538	0,1809	G	22,44	0,00000217	8,5560
2	rs12712049	99575262	A	0,0000	0,4634	G	22,24	0,00000240	0,0000
8	rs2920922	76870632	C	0,2500	0,7931	T	21,58	0,00000339	0,0870
18	rs6566904	53662113	T	0,3846	0,0455	A	21,43	0,00000367	13,1200
4	rs1395612	189670284	A	0,1923	0,7051	G	21,01	0,00000457	0,0996
11	rs11019672	91256822	A	0,5714	0,1444	G	21,00	0,00000458	7,8970
1	rs354636	91245081	G	0,1250	0,6346	A	20,87	0,00000490	0,0823
2	rs1430804	15112010	G	0,5385	0,1081	A	20,77	0,00000519	9,6250
2	rs17417862	54973424	G	0,1538	0,6667	A	20,63	0,00000556	0,0909
17	rs12600420	32122480	C	0,0000	0,5714	T	20,57	0,00000575	0,0000
4	rs1390911	44627008	C	0,2692	0,7833	T	20,52	0,00000590	0,1019
9	rs1335445	11349686	G	0,0714	0,5641	T	20,36	0,00000642	0,0594
6	rs2788669	120978396	C	0,1538	0,6585	T	20,22	0,00000689	0,0943
13	rs7997862	41469136	G	0,2143	0,7000	A	20,01	0,00000770	0,1169

A correção de Bonferroni foi aplicada obtendo-se P significativa ($P=0,015$) somente para o SNP rs17469997 do cromossomo 10. Segue abaixo a Tabela 4 com os valores de P após a correção de Bonferroni. onde CRM = cromossomo, # SNP = número de referência do SNP, POSIÇÃO = referente à posição do SNP no cromossomo em questão, P = probabilidade das diferenças entre caso e controle ser devida ao acaso e P BONFERRONI = valores de P após a correção de Bonferroni.

Tabela 4 – Valores de P para os 18 SNPs após a correção de Bonferroni.

CRM	# SNP	POSIÇÃO	P	P BONFERRONI
10	rs17469997	8495938	0,00000006	0,0157
7	rs1568799	42634401	0,00000043	0,1119
11	rs4923076	23203347	0,00000187	0,4883
X	rs4827928	97413650	0,00000188	0,4908
4	rs2086048	54221234	0,00000217	0,5671
2	rs12712049	99575262	0,00000240	0,6280
8	rs2920922	76870632	0,00000339	0,8866
18	rs6566904	53662113	0,00000367	0,9592
4	rs1395612	189670284	0,00000457	~ 1,0000
11	rs11019672	91256822	0,00000458	~ 1,0000
1	rs354636	91245081	0,00000490	~ 1,0000
2	rs1430804	15112010	0,00000519	~ 1,0000
2	rs17417862	54973424	0,00000556	~ 1,0000
17	rs12600420	32122480	0,00000575	~ 1,0000
4	rs1390911	44627008	0,00000590	~ 1,0000
9	rs1335445	11349686	0,00000642	~ 1,0000
6	rs2788669	120978396	0,00000689	~ 1,0000
13	rs7997862	41469136	0,00000770	~ 1,0000

Um estudo de mistura étnica foi realizado entre os grupos caso e controle através do programa GENIOC (CABELLO e KRIEGER, 1997). No grupo caso, foi observada uma maior proporção da etnia africana com 37,59%, com 35,03% de etnia ameríndia e 27,38% de etnia caucasiana (Tabela 5).

Tabela 5 – Mistura racial dos indivíduos do Grupo Caso.

GRUPO CASO	PROPORÇÃO	ERRO PADRÃO
AFRICANOS	0,3759	0,4159
AMERÍNDIOS	0,3503	0,2978
CAUCASIANOS	0,2738	0,3411

No grupo controle a maior proporção foi observada na etnia caucasiana com 45,79%, com 39,35% da etnia ameríndia e 14,86% da etnia africana (Tabela 6).

Tabela 6 – Mistura racial dos indivíduos do Grupo Controle.

GRUPO CONTROLE	PROPORÇÃO	ERRO PADRÃO
AFRICANOS	0,1486	0,3892
AMERÍNDIOS	0,3935	0,2540
CAUCASIANOS	0,4579	0,3361

Os testes de equilíbrio de Hardy-Weinberg também foram realizados para os 18 SNPs mais significativos do teste de associação com o programa PLINK. A seguir na Tabela 7 temos os testes de equilíbrio de Hardy-Weinberg para cada um

dos SNPs dos indivíduos afetados divididos em três tipos de teste cada. Onde CRM = cromossomo, # SNP = número de referência do SNP, TESTE = categoria do indivíduo (ALL = todos indivíduos da amostra, AFF = indivíduos afetados da amostra e UNAF = indivíduos não-afetados da amostra), GENÓTIPO = frequência de cada genótipo e P = probabilidade do teste.

Tabela 7 – Tabela dos testes de equilíbrio de Hardy-Weinberg para os 18 SNPs mais significativos.

CRM	# SNP	TESTE	A1	A2	GENÓTIPO	P
10	rs17469997	ALL	C	G	3/2/19	0,00453
10	rs17469997	AFF	C	G	3/2/9	0,02609
10	rs17469997	UNAFF	C	G	0/0/10	1,00000
7	rs1568799	ALL	A	C	10/2/12	0,00003
7	rs1568799	AFF	A	C	3/2/10	0,02076
7	rs1568799	UNAFF	A	C	7/0/2	0,01176
11	rs4923076	ALL	T	A	6/6/12	0,02855
11	rs4923076	AFF	T	A	6/3/4	0,08553
11	rs4923076	UNAFF	T	A	0/3/8	1,00000
X	rs4827928	ALL	A	G	4/2/6	0,02867
X	rs4827928	AFF	A	G	4/0/0	1,00000
X	rs4827928	UNAFF	A	G	0/2/6	1,00000
4	rs2086048	ALL	A	G	6/7/8	0,18470
4	rs2086048	AFF	A	G	5/5/2	1,00000
4	rs2086048	UNAFF	A	G	1/2/6	0,34120
2	rs12712049	ALL	A	G	2/2/17	0,02764
2	rs12712049	AFF	A	G	0/0/14	1,00000
2	rs12712049	UNAFF	A	G	2/2/3	0,44060
8	rs2920922	ALL	C	T	6/5/10	0,02491
8	rs2920922	AFF	C	T	1/4/7	1,00000
8	rs2920922	UNAFF	C	T	5/1/3	0,03167
18	rs6566904	ALL	T	A	3/5/15	0,07110
18	rs6566904	AFF	T	A	3/4/5	0,28930
18	rs6566904	UNAFF	T	A	0/1/10	1,00000

(continua)

(continuação)

CRM	# SNP	TESTE	A1	A2	GENÓTIPO	P
4	rs1395612	ALL	A	G	6/6/9	0,07443
4	rs1395612	AFF	A	G	0/3/9	1,00000
4	rs1395612	UNAFF	A	G	6/3/0	1,00000
11	rs11019672	ALL	A	G	5/7/10	0,17360
11	rs11019672	AFF	A	G	5/4/3	0,28930
11	rs11019672	UNAFF	A	G	0/3/7	1,00000
1	rs354636	ALL	G	A	4/3/14	0,00776
1	rs354636	AFF	G	A	0/3/11	1,00000
1	rs354636	UNAFF	G	A	4/0/3	0,01166
2	rs1430804	ALL	G	A	4/5/11	0,11380
2	rs1430804	AFF	G	A	4/4/4	0,28150
2	rs1430804	UNAFF	G	A	0/1/7	1,00000
2	rs17417862	ALL	G	A	3/5/13	0,09295
2	rs17417862	AFF	G	A	0/2/10	1,00000
2	rs17417862	UNAFF	G	A	3/3/3	0,50240
17	rs12600420	ALL	C	T	5/3/10	0,00868
17	rs12600420	AFF	C	T	0/0/8	1,00000
17	rs12600420	UNAFF	C	T	5/3/2	0,47990
4	rs1390911	ALL	C	T	7/7/8	0,09892
4	rs1390911	AFF	C	T	1/5/7	1,00000
4	rs1390911	UNAFF	C	T	6/2/1	0,34120
9	rs1335445	ALL	G	T	3/6/13	0,27000
9	rs1335445	AFF	G	T	0/2/11	1,00000
9	rs1335445	UNAFF	G	T	3/4/2	1,00000
6	rs2788669	ALL	C	T	5/2/14	0,00089
6	rs2788669	AFF	C	T	1/1/10	0,13040
6	rs2788669	UNAFF	C	T	4/1/4	0,03645
13	rs7997862	ALL	G	A	5/7/11	0,17140
13	rs7997862	AFF	G	A	3/0/11	0,00097
13	rs7997862	UNAFF	G	A	2/7/0	0,17650

Estes testes de equilíbrio de Hardy-Weinberg mostraram que os SNPs rs1430804 e rs17417862 no cromossomo 2, rs1390911, rs1395612 e rs2086048 no cromossomo 4, rs1335445 no cromossomo 9, rs11019672 no cromossomo 11, rs7997862 no cromossomo 13 e rs6566904 no cromossomo 18 se apresentaram em equilíbrio mostrando valores de P superiores a 0,05 para a amostra total.

Enquanto que para os indivíduos afetados para a infecção por *Trypanosoma cruzi* nos SNPs rs1568799 no cromossomo 7, rs17469997 no cromossomo 10 não se apresentaram em equilíbrio de Hardy-Weinberg, assim como nos indivíduos não afetados para os SNPs rs354636 no cromossomo 1, rs2788669 no cromossomo 6, rs1568799 no cromossomo 7 e rs2920922 no cromossomo 8.

A seguir, testes de ligação gênica foram realizados com o programa SuperLink (FISHELSON e GEIGER, 2002), que faz testes de ligação do tipo dois-pontos. Foram analisados todos os cromossomos com um espaçamento de 50 centimorgans com 1440 SNPs informativos, obtendo valores de *lod score* iguais a zero (figura 7).

Um segundo teste de ligação foi realizado com espaçamento entre os SNPs diminuído para 25 cM com 1440 SNPs informativos, o melhor resultado de *lod score* obtido foi de 0,2271 no SNPA1949991 rs4887855 no cromossomo 16 (Figura 8).

Com o programa GeneHunter (KRUGLYAK et al., 1996), que faz testes de ligação do tipo multi-ponto, foram analisados todos os cromossomos com espaçamento de 50 centiMorgans (cM), no total foram 1440 SNPs informativos. O

maior resultado apresentou *lod score* igual a 1,6521 no cromossomo 18 SNPA4241293 rs2712747 (Figura 9).

Com base nos resultados de ligação do cromossomo 18, um novo teste multi-ponto foi realizado neste cromossomo com espaçamento de 20 cM, resultando no valor máximo de *lod score* igual a 1,5052 (Figura 10).

O último teste foi realizado no cromossomo 10, que havia apresentado associação com a infecção por *T. cruzi*, a ligação multi-ponto apresentou resultado máximo de *lod score*, com espaçamento de 20 cM, um valor igual a 0,6021 (Figura 11).

Figura 7 – Ligação dois-pontos em todos os cromossomos com espaçamento de 50cM.

Project:	EASYLINKAGE	Inheritance:	Dominant	Marker	CHR	cM	LOD	Theta
Family name:	TOTALS	Common allele:	99.90 %	1. SNPA2149899	1	9.22	0.0000	0.0000
Used map:	AFFY 500k Marshfield 06042007 v3 (sex)	Dominant allele:	0.10 %	2. SNPA2214366	1	60.46	0.0000	0.0000
Marker positions:	75 ok / 2 ? / 0 outside	LC1 PCOPY rate:	0.00 %	3. SNPA2059879	1	110.92	0.0000	0.0000
Allele frequencies:	All individuals from marker file	LC1 PENET wt/mt:	100.00 %	4. SNPA1836676	1	165.08	0.0000	0.0000
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %	5. SNPA2085365	1	216.67	0.0000	0.0000

Analysis time: 15.02.2011 - 14:08:28
Time elapsed: 23 s
Path: C:\Program Files\easyLINKAGE
Pedigree structure file: pedhosp3.pro - SNP data file: cmmalho3.snp

Marker coverage and chromosomes

Copyright ©, Tom H. Lindner & Katrin Hoffmann
easyLINKAGE Plus v3.08, January 12, 2007

Figura 8 – Ligao dois-pontos em todos os cromossomos com espaamento de 25cM.

Project:	EASYLINKAGE	Inheritance:	Dominant	Marker	CHR	cM	LOD	Theta
Family name:	TOTALS	Common allele:	99.90 %	1. SNPA1949991	16	91.59	0.2271	0.1500
Used map:	AFFY 500k Marshfield 06042007 v3 (sex)	Minor allele:	0.10 %	2. SNPA4228972	20	37.19	0.1242	0.2000
Marker positions:	133 ok / 2 ? / 0 outside	LC1 PCOPY rate:	0.00 %	3. SNPA2149899	1	9.22	0.0000	0.0000
Allele frequencies:	All individuals from marker file	LC1 PENET wt/int:	100.00 %	4. SNPA2131331	1	34.30	0.0000	0.0000
CALC interval:	Entire chromosome	LC1 PENET int/int:	100.00 %	5. SNPA2214366	1	60.46	0.0000	0.0000

Analysis time: 15.02.2011 – 14:05:16
Time elapsed: 41 s
Path: C:\Program Files\easyLINKAGE
Pedigree structure file: pedhosp3.pro – SNP data file: crnallhosp.snp

Marker coverage and chromosomes

Copyright , Tom H. Lindner & Katrin Hoffmann
easyLINKAGE Plus v5.08, January 12, 2007

Figura 9 – Ligação multi-ponto em todos os cromossomos com espaçamento de 50cM.

Project:	EASYLINKAGE	Inheritance:	Dominant	Marker	CHR	cM	pLOD	NPL	npLOD	P val	INFO	Alpha	hLOD
Family name:	TOTALS	Common allele:	99.90 %	1. SNPA4241293	18	1.88	1.5052	0.0000	0.0000	1.0000	0.3767	1.0000	1.5051
Used map:	AFFY 500k Marshfield 06042007 v3 (see Dismiss)	Dismissing allele:	0.10 %	2. SNPA1878453 – SNPA4230005	18	7.33	1.3451	0.0000	0.0000	1.0000	0.2275	1.0000	1.3451
Marker positions:	6 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %	3. SNPA1878453 – SNPA4230005	18	12.79	1.1967	0.0000	0.0000	1.0000	0.1565	1.0000	1.1967
Allele frequencies:	All individuals from marker file	LC1 PENET wt/mt:	100.00 %	4. SNPA1878453 – SNPA4230005	18	18.24	1.0595	0.0000	0.0000	1.0000	0.1155	1.0000	1.0595
HAP algo / SCFet:	Off / All	LC1 PENET mt/mt:	100.00 %	5. SNPA2264383 – SNPA4230476	18	23.70	0.9334	0.0000	0.0000	1.0000	0.0930	1.0000	0.9334

Pair setting: All pairs of affected / phenotyped sibs

GeneHunter v2.1r5 (MultiPoint) – Computation in sets of 50 markers, spacing 20.0000 cM

Analysis time: 15.02.2011 – 14:25:36
Time elapsed: 1 s
Path: C:\Program Files\easyLINKAGE
Pedigree structure file: pedhoop3.pro – SNP data file: crn18hoop.snp

Copyright ©, Tom H. Lindner & Katrin Hoffmann
easyLINKAGE Plus v5.08, January 12, 2007

Figura 10 – Ligação multi-ponto no cromossomo 18 com espaçamento de 20cM.

Project:	EASYLINKAGE	Inheritance:	Dominant	Marker	CHR	cM	pLOD	NPL	apLOD	P val	INFO	Alpha	hLOD
Family name:	TOTALS	Common allele:	99.90 %	1. SNPA4241293	18	1.88	1.6521	0.0000	0.0000	1.0000	0.5030	1.0000	1.6521
Used map:	AFFY 500k Marshfield 06042007 v3 (see Dismap)	Rare allele:	0.10 %	2. SNPA1878453 – SNPA4230005	18	11.92	1.4240	0.0000	0.0000	1.0000	0.1963	1.0000	1.4240
Marker positions:	75 ok / 2 ? / 0 outside	LC1 PCOPY rate:	0.00 %	3. SNPA2264383 – SNPA4230476	18	21.97	1.2508	0.0000	0.0000	1.0000	0.0990	1.0000	1.2508
Allele frequencies:	All individuals from marker file	LC1 PENET wt/mt:	100.00 %	4. SNPA4230476 – SNPA1937045	18	32.01	1.1316	0.0000	0.0000	1.0000	0.0821	1.0000	1.1316
HAP algo / SCFct:	OH / All	LC1 PENET mt/mt:	100.00 %	5. SNPA1937045 – SNPA2312714	18	42.05	1.0636	0.0000	0.0000	1.0000	0.1351	1.0000	1.0636

Pair setting: All pairs of affected / phenotyped sibs

pLOD (MPT) GeneHunter v2.1r5 (MultiPoint) – Computation in sets of 50 markers, spacing 50.0000 cM

Analysis time: 15.02.2011 – 13:42:21
Time elapsed: 6 s
Path: C:\Program Files\easyLINKAGE
Pedigree structure file: pedhosp3.pro – SNP data file: crmallhosp.snp

Marker coverage and chromosomes

Copyright ©, Tom H. Lindner & Katrin Hoffmann
easyLINKAGE Plus v5.08, January 12, 2007

Figura 11 – Ligação multi-ponto no cromossomo 10 com espaçamento de 20cM.

Project:	EASYLINKAGE	Inheritance:	Dominant	Marker	CHR	cM	pLOD	NPL	npLOD	P val	INFO	Alpha	hLOD
Family name:	TOTALS	Common allele:	99.90 %	1. SNPA4230449	10	46.81	0.6021	0.0000	0.0000	1.0000	1.0000	1.0000	0.6021
Used map:	AFFY 500k Marshfield 06042007 v3 (sex)	Discarded allele:	0.10 %	2. SNPA2278369 – SNPA4205064	10	50.82	0.5580	0.0000	0.0000	1.0000	0.7272	1.0000	0.5580
Marker positions:	7 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %	3. SNPA2294570 – SNPA4230449	10	42.22	0.5390	0.0000	0.0000	1.0000	0.6366	1.0000	0.5390
Allele frequencies:	All individuals from marker file	LC1 PENET wt/mt:	100.00 %	4. SNPA2286341 – SNPA4226323	10	54.83	0.5220	0.0000	0.0000	1.0000	0.5993	1.0000	0.5220
HAP algo / SCFet:	Off / All	LC1 PENET mt/mt:	100.00 %	5. SNPA4226323 – SNPA1884460	10	58.84	0.4940	0.0000	0.0000	1.0000	0.5360	1.0000	0.4940

Pair setting: All pairs of affected / phenotyped sibs

GeneHunter v2.1r5 (MultiPoint) – Computation in sets of 50 markers, spacing 20.0000 cM

Analysis time: 15.02.2011 – 14:01:46
Time elapsed:
Path: C:\Program Files\easyLINKAGE
Pedigree structure file: pedhosp3.pro – SNP data file: cml10hosp.snp

Copyright ©, Tom H. Lindner & Katrin Hoffmann
easyLINKAGE Plus v5.08, January 12, 2007

6 DISCUSSÃO

A amostra estudada proveniente da Hospedaria de Imigrantes do Departamento de Colonização do Estado de São Paulo e composta por 69 indivíduos divididos em 14 famílias nucleares, selecionados por apresentarem ao menos um dos pais infectado pelo *Trypanosoma cruzi*.

O DNA foi extraído das amostras de células vermelhas que foram centrifugadas e separadas do soro, onde se trabalhava com a hipótese de se encontrar algum vestígio de células nucleadas nessas amostras de hemácias em glicerol.

Sabe-se que nas amostras de sangue, o DNA se degrada com o passar do tempo, perdendo quantidade e qualidade mesmo estando estocado corretamente em freezer -80°C e, levando em conta os 40 anos de idade da amostra de sangue utilizada neste trabalho, era de se suspeitar que as dificuldades na extração de DNA passível de análises dessas amostras, não seriam desprezíveis (ROSS; HAITES; KELLY, 1990).

A qualidade e quantidade de DNA foram muito prejudicadas devido a pouca quantidade de células nucleadas e devido a fatores como: tempo de coleta da amostra superior a 40 anos, realizada entre 1969 e 1970, transporte para outros laboratórios de universidade e institutos nos quais o Professor Doutor Henrique Krieger veio a trabalhar e problemas com o armazenamento devido à quebra do freezer -80°C onde as amostras estavam estocadas.

Deve-se enfatizar que existem vários protocolos e kits desenvolvidos para a extração de DNA de acordo com a sua fonte, seja ela sangue, saliva ou qualquer outro tipo de material. Após várias tentativas de obtenção de DNA dessa amostra, esses métodos comuns de extração de DNA como extração por coluna e solução salina se mostraram pouco eficientes. Outros métodos como o “magnetic beads” onde a extração é feita com diferenças de cargas positivas e negativas também não obtiveram sucesso.

Para obter êxito na extração do DNA foram necessárias, algumas modificações na técnica de extração de solução salina com a técnica de fenol/clorofórmio e purificação com glicogênio e álcool.

Tendo em vista que os aspectos técnicos deste trabalho são de importância capital para uma apreciação de sua validade e também pelo seu papel no conjunto de abordagens que viriam a seguir, passamos a discutir esses aspectos em algum detalhamento.

Hoje uma plataforma de microarranjos de DNA é capaz de analisar mais de um milhão de SNPs de uma única amostra de DNA, a qual, obviamente, requer certa quantidade de DNA em boa qualidade.

Dessa forma, foi necessário aumentar a quantidade de DNA das amostras, uma vez que foi verificado que a quantidade e a concentração não atingiram 250 ng, portanto, não sendo suficientes para a genotipagem com os chips de DNA pretendidos. Utilizou-se então a enzima chamada Phi 29 (Illustra GenomiPhi™ DNA Amplification Kit), que amplifica todo DNA presente na amostra.

Para comprovar se o DNA extraído e amplificado pela enzima Phi 29, era realmente proveniente da amostra dos indivíduos da Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo e não de uma possível contaminação da amostra por fungos ou bactérias, foi realizada uma reação de PCR (Reação de polimerase em cadeia) ao acaso utilizando microsátélites ou STR (*Short Tandem Repeats*).

Os primers usados nessa reação de PCR são especificamente humanos, de forma que a visualização das bandas de DNA em um gel de agarose comprova que o DNA extraído e amplificado com Phi 29 é proveniente da amostra da Hospedaria de Imigrantes do Departamento de Imigração e Colonização em São Paulo e não de uma contaminação por fungos e bactérias (Figura 4 e 5).

A seguir, cada amostra de DNA foi quantificada utilizando o equipamento NanoDrop™ 2000 (NanoDrop Technologies) (Tabela 2) e tiveram sua concentração final ajustada para 50 ng/μl, permitindo dar início a varredura genômica utilizando a plataforma GeneChip System Affymetrix com o microarranjo Genechip Mapping 250K Nsp I, capaz de fornecer genótipos de 260 K SNPs. Posteriormente foram realizadas as reações de digestão, ligação, diluição, PCR, purificação, eluição, normalização, fragmentação, marcação, hibridação, lavagem e escaneamento.

Após a genotipagem, os dados foram preparados para permitir a análise através de programas de computador. Uma averiguação de legitimidade é feita automaticamente antes dos testes de associação pelo programa PLINK (Figura 6), onde qualquer indício de ilegitimidade culmina na eliminação do indivíduo

em questão. Tendo em vista que não houve evidência de ilegitimidade na amostra utilizada, descartamos a hipótese de contaminação cruzada.

Os testes de associação revelaram diferença significativa entre casos e controles, com probabilidade de 0,0157 depois da correção de Bonferroni para o SNP rs17469997 no cromossomo 10 (Tabela 4).

Essa correção é importante pois leva em conta que a probabilidade de se encontrar, com os níveis de significância habituais, um dado significativo ao acaso entre 260 mil tentativas acaba se tornando muito grande. Sendo assim, aplicando a fórmula de correção tem-se o valor de P ajustado para estimativas individuais (BLAND, 1995). O SNP rs17469997 localizado no cromossomo 10, está localizado a 84 kb do STR (*short tandem repeat*) D10S1459 e localizado a 339 kb do gene GATA3.

Não conhecemos (até o mês de outubro) referências a respeito do STR D10S1459.

O gene GATA 3, codifica a proteína que pertence a família GATA de fatores de transcrição, é um regulador do desenvolvimento de células T e tem papel importante na biologia de células endoteliais. Estes fatores de transcrição são responsáveis pela diferenciação celular e mutações nesses fatores de transcrição e estão associados com doenças genéticas devido a falhas na diferenciação celular específica para linhagens de células T (MILLER, 2009).

O GATA-3 é necessário para o desenvolvimento de célula pré-T que dão origem aos linfócitos T (HOSOYA, 2009), além de estar associado com algumas doenças como eczema, (ARSHAD, 2008), câncer de mama

(KOUROS-MEHR, 2008; GARCIA-CLOSAS, 2007) hipoparatiroidismo, surdez e síndrome de displasia renal (OMIN, 2011; ZAHIRIEH, 2005; MUROUYA, 2001).

Associações com outras infecções intra-celulares não foram encontradas no presente trabalho, como: a malária, onde Sikora (2011) e Driss (2011) encontraram associação no cromossomo 12, Jallow (2009) encontrou associação no cromossomo 11 e Beiguelman (2003) no cromossomo 1. A hepatite, onde foram encontradas associações com os cromossomos 19, 16 e 6 (SUGIYAMA e MIZOKAMI, 2011, ASSELAH, 2011, LIU, 2011, HOSHIDA, 2011). Hanseníase, encontrada associação com o cromossomo 6 (BAKIJA-KONSUO, 2011). E a leishmaniose, associada ao cromossomo 6 (FAKIOLA, 2011) e cromossomo 5 (JERONIMO, 2007).

Ainda que somente um SNP (rs17469997) mostrou associação significativa para a infecção chagásica, os outros 17 SNPs restantes não devem ser descartados pois podem ser indicativos de locais próximos a genes relacionados com a resistência/suscetibilidade a infecção por *T. cruzi*.

Na tentativa de comprovar os estudos de associação e evitar problemas de estratificação, um teste final de ligação gênica foi realizado com o programa GeneHunter (KRUGLYAK et al., 1996) para a ligação multi-ponto e o programa SuperLink (FISHELSON e GEIGER, 2002) para a ligação dois-pontos.

No estudo de ligação é verificada a taxa de recombinação dos alelos, sendo que genes próximos uns aos outros têm uma chance menor de recombinação e, dessa forma, genes mais distantes têm uma maior chance de

recombinação. Espera-se que a taxa de recombinação dentro dos indivíduos de uma mesma família mantenha-se constante no decorrer das gerações.

Com dados familiares e com a genotipagem desses indivíduos é possível obter a taxa de recombinação entre marcadores e o fenótipo infectado ou não infectado pelo *Trypanosoma cruzi*.

Na análise de ligação foi utilizado o método *lod score* (*logarithm of odds ratio*), onde *odd ratio* é definido como a razão entre a probabilidade de um genótipo composto específico ser observado em determinada família, em uma determinada fração de recombinação diferente de 0,5 contra a mesma probabilidade, computada considerando segregação independente.

Quanto maior o valor de *lod score*, maior a evidência a favor de ligação, dessa forma, um valor de *lod score* maior que 3 é altamente indicativo de ligação (MORTON, 1955).

As análises de ligação foram realizadas com 14 famílias com o intuito de verificar se existe alguma ligação entre os genes e relacioná-los com a infecção por *Trypanosoma cruzi*. Os programas realizam automaticamente uma averiguação dos dados dentro das famílias. Dessa forma, alguns indivíduos acabaram não sendo utilizados nos testes de ligação devido á ilegitimidade, assim como uma família inteira foi excluída por não apresentar pelo menos um dos pais infectados pelo *Trypanossoma cruzi*, critério fundamental para a inclusão na amostra.

O primeiro teste de ligação foi realizado com o programa SuperLink (FISHELSON e GEIGER, 2002), que faz testes de ligação do tipo dois-pontos.

Foram analisados todos os cromossomos com um espaçamento de 50 centimorgans com 1440 SNPs informativos, mas não foram obtidos resultados significativos e os valores de lod score foram iguais a zero (Figura 7). Devido a esses resultados iniciais, um segundo teste de ligação foi realizado com espaçamento entre os SNPs diminuído para 25 cM com 1440 SNPs informativos, o melhor resultado foi o SNPA1949991 rs4887855 no cromossomo 16 que está associado com doenças como diabetes, esclerose múltipla (MARTÍNEZ, 2010) autismo (KUMAR, et al., 2009), esquizofrenia (KUROTAKI, et al., 2011), degeneração macular (SPENCER, et al., 2011) e obesidade (WALTERS, et al., 2010), porém o *lod score* obtido foi de 0,2271, sendo dessa forma não significativo (Figura 8).

Com o programa GeneHunter (KRUGLYAK et al., 1996), que faz testes de ligação do tipo multi-ponto, foram analisados todos os cromossomos com espaçamento de 50 centimorgans (cM), no total foram 1440 SNPs informativos. O melhor resultado foi no cromossomo 18 no SNPA4241293 rs2712747 (Figura 9). Este cromossomo está relacionado com a doença de Parkinson (GAO, et al., 2009), tuberculose (THYE, et al., 2010), doenças no fígado (CHALASANI, et al., 2010) e dislexia (SCERRI, et al., 2010). No entanto, apresentou *lod score* igual a 1,6521 não se mostrando significativo.

Com base nos resultados de ligação do cromossomo 18, um novo teste multi-ponto foi realizado neste cromossomo com espaçamento de 20 cM, na tentativa de se encontrar significância através de valores de lod score maiores que 3. O resultado do programa GeneHunter (KRUGLYAK et al., 1996)

apresentou valor máximo de lod score igual a 1,5052, sendo assim novamente não significativa (Figura 10).

O último teste foi realizado no cromossomo 10 (já que havia apresentado associação com a infecção por *T. cruzi*) com o programa GeneHunter (KRUGLYAK et al., 1996) com teste de ligação multi-ponto. Os resultados de lod score com espaçamento de 20 cM apresentaram maior valor igual a 0,6021 não apresentando significância (Figura 10).

Por outro lado, este trabalho não corrobora as evidências de associações de SNPs e a infecção por *T. cruzi* encontradas no SNP rs2430561, localizado no cromossomo 12, alelo +874T/A do gene INFG (interferon gama) em uma população da Colômbia (TORRES et al, 2010); SNP rs755622, no cromossomo 22, alelo -173G/C do gene MIF (fator inibitório de migração de macrófagos) em populações da Colômbia e Peru (TORRES et al, 2009) e SNPs no gene TGF β 1 (fator de transformação do crescimento), localizado no cromossomo 19 em populações da Colômbia e Peru (CALZADA et al, 2009).

Outras associações relacionadas especificamente com a cardiopatia conseqüente da infecção por *T. cruzi* também não foram reproduzidas no presente estudo, algumas delas são: o gene *TNF* (fator de necrose tumoral), localizado no cromossomo 6, também em uma população brasileira (DRIGO et al, 2006); o SNPs rs2239704, alelo +80A/C e o SNP rs909253, alelo +252A/G do gene *LTA* (linfotóxina), no cromossomo 6 em uma população do Brasil (RAMASAWMY et al, 2007) e os genes *TLR* e *MAL/TIRAP* (*toll like receptor* e

toll interleukin receptor adaptor protein), localizados nos cromossomos 1, 3, 4, 9 e 11, em uma população brasileira (RAMASAWMY et al, 2009).

Deve ser salientado, que mesmo os testes realizados após a aplicação de alguns filtros e dessa forma, utilizando apenas os SNPs que foram genotipados em todos os indivíduos de modo a obter a melhor informação sobre os mesmos, não alteraram os resultados que incluíram toda a amostra.

Novos trabalhos devem ser realizados não apenas com a intenção de aumentar o tamanho desta amostra, mas também buscar novos indícios de associação (meta-análise, por exemplo) que possam ser comprovados com estudos de ligação gênica e testar definitivamente a associação encontrada no cromossomo 10. A confirmação desses resultados poderia contribuir tanto para a prevenção como para a utilização de novas terapias para esta infecção.

7 CONCLUSÕES

1. Neste trabalho foram obtidos resultados positivos na extração de DNA em amostras antigas e mal conservadas. As técnicas foram descritas em detalhes que serão úteis para estudos semelhantes.
2. Testes de GWAS indicaram associação entre uma região do cromossomo 10 e a infecção por *T. cruzi* (SNP rs17469997).
3. Outros resultados presentes na literatura, tanto na infecção por *T. cruzi*, como formas da doença não foram confirmados no presente estudo.
4. Alguns estudos que mostraram associação com outras infecções intracelulares não encontraram equivalência no presente estudo.

REFERÊNCIAS

ALLISON, A. C.; PHILL. D. Protection afforded by sickle-cell trait against subtertian malarial infection. **Br. Med. J.**, v. 2, p. 290-294, 1954.

ASSELAH, T., et al. IL28B polymorphism is associated with treatment response in patients with genotype 4 chronic hepatitis C. **J. Hepatol.** Disponível em <<http://www.ncbi.nlm.nih.gov/pubmed/21951981>>, Acesso em: 02 Nov. 2011.

BAKIJA-KONSUO, A., et al. Leprosy epidemics during history increased protective allele frequency of PAK2/PACRG genes in the population of the Mijet Island, Croatia. **Eur. J. Med. Genet.**, v. 6, p. e548-552, 2011.

BARBOSA, C. A. A. **Aspectos médico-sociais de uma população rural brasileira.** 1971. 75 f. Dissertação (Mestrado em Ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1971.

_____.; KOURY, W. H.; KRIEGER, H. Linkage data on MN and the Hb β locus. **Ame. J. Hum. Genet.**, v. 27, p. 797-801, 1975.

_____. **Estudo genético-epidemiológico de algumas variáveis hemodinâmicas.** 1976. 130 f. Tese (Doutorado em Ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1976.

BEIGUELMAN, B., et al. The association of genetic markers and malaria infection in the Brazilian Western Amazonian Region. **Mem. Inst. Oswaldo Cruz.**, v. 98, p. 455-460, 2003.

_____. **Dinâmica dos genes nas famílias e nas populações.** 2. ed. Ribeirão Preto: Sociedade Brasileira de Genética, 1994. 460p.

_____. Some remarks on the genetics of leprosy resistance. **Acta. Genet. Méd. Gemellol.**, v. 17, p. 584-594, 1968.

_____. The genetics of resistance to leprosy. **Int. J. Lepr.**, v. 33 , p. 808-812, 1965.

BELL, J. Predicting disease using genomics. **Nature**, v. 429, p.453-456, 2004.

BELLAMY, R. et al. Genetic susceptibility to tuberculosis in Africans: A genome-wide scan. **Proc. Nat. Acad. Sci.**, v. 97, p. 8005-8009, 2000.

BLAND, J. M.; ALTMAN, D. G. Mutliple significance tests: the Bonferroni method. **Br Med J**, v. 310, p.170, 1995.

De acordo com:

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023:** informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

BRACKEN, M. B. Genomic epidemiology of complex disease: the need for an electronic evidence-based approach to research synthesis. **Am J Epidemiol.**, v. 162, n. 4, p. 297-301, 2005.

CABELLO, P. H. A. **Freqüências gênicas e coeficiente de endocruzamento em uma população rural brasileira.** 1972. 90 f. Dissertação (Mestrado em Ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1972.

_____. **Caracterização demográfica e estrutura genética de uma população migrante brasileira.** 1976. 127 f. Tese (Doutorado em Ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1976.

_____; KRIEGER, H. Genioc: sistema para análises de dados de genética. Rio de Janeiro: FIOCRUZ, 1997. 139 p.

CALZADA, J. E.; BERAÚN, Y.; GONZÁLEZ, C. I.; MARTÍN, J. Transforming growth factor beta 1 (TGF β 1) gene polymorphism and Chagas disease susceptibility in Peruvian and Colombian patients. **Cytokine**, v. 45, p. 149-153, 2009.

CAMARGO, E. P. Growth and differentiation in *Trypanosoma cruzi*. I. Origin of metacyclic trypanosomes in liquid media. **Rev. Inst. Med. Trop. São Paulo**, v. 6, p. 93-100, 1964.

CARRINGTON, M., et al. HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. **Science**, v. 283, p. 1748-1752, 1999.

CHAKRAVARTI, M. R.; VOGEL, F. **A twin study on leprosy.** *Stuttgart: Thieme*, 1973. 124p.

CHALAZANI, N., Genome-wide association study identifies variants associated with histologic features of nonalcoholic fatty liver disease. **Gastroent.**, v. 139, p. 1567-1576, 2010.

COMSTOCK, G. W. Tuberculosis in twins: a re-analysis of the Proffit survey. **Am. Rev. Respir. Dis.**, v. 117, p. 621-624, 1978.

DAL COLLETO, G. M. D. **Taxa de ilegitimidade de uma população rural brasileira.** 1972. 67 f. Tese (Doutorado em ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1972.

_____; KRIEGER, H.; MAGALHÃES, J. R. Estimates of the genetical and environmental determinants of serum lipid and lipoprotein concentration in brazilian twins. **Hum. Hered.**, v. 31, p. 232-237, 1981.

DRIGO, S. et al. TNF gene polymorphism are associated with reduced survival in severe Chagas' disease cardiomyopathy patients. **Microbes Infect.**, v. 8, p. 598-603, 2006.

DRISS, A., et al. Genetic polymorphisms linked to susceptibility to malaria. **Mal. J.**, v. 10, p. 271-281, 2011.

FAKIOLA, M. et al. Genetic and functional evidence implicating DLL1 as the gene that influences susceptibility to visceral leishmaniasis at chromosome 6q27. **J. Infect. Dis.**, v. 3, p. 467-477, 2011.

FEITOSA, M. F. **Aspectos genético-epidemiológicos da infecção chagásica em nordestinos brasileiros**. 1985. 185 f. Dissertação (Mestrado em ciências) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 1985.

_____; KRIEGER, H. An appraisal of the epidemiology of *Trypanosoma cruzi* serology in Brazil. **Mem. Inst. Oswaldo Cruz**, v. 86, p. 159-167, 1991.

_____, et al. The genetic epidemiology of leprosy in a Brazilian population. **Am. J. Hum. Genet.**, v. 56, p. 1179-1185, 1995.

FISHELSON, M; GEIGER, D. Exact linkage computations for general pedigrees. **Bioinformatics**, v. 18, p. 189-198, 2002

FITNESS, J. et al. Large-scale candidate gene study of leprosy susceptibility in the karonga district of Northern Malawi. **Am. J. Trop. Med. Hyg.**, v. 71, p. 330-340, 2004.

FLORI, L., et al. Linkage of mild malaria to the major histocompatibility complex in families living in Burkina Faso. **Hum. Mol. Genet.**, v. 12, p. 375-378, 2003.

FRAZER, A. K.; MURRAY, S. S.; SCHORK, J. N.; TOPOL, J. E. Human genetic variation and its contribution to complex traits. **Nat. Genet.**, v. 10, p. 241-251, 2009.

FUNDAÇÃO OSWALDO CRUZ (FIOCRUZ). Doença de Chagas. Disponível em: <<http://www.fiocruz.br/chagas/cgi/cgilua.exe/sys/start.htm?sid=34>>. Acesso em: 24 jan. 2011.

GAO, X., et al. Genome-wide linkage screen in familial Parkinson disease identifies loci on chromosomes 3 and 18. **Am. J. Hum. Genet.**, v. 84, p. 499-504, 2009.

GOLDSTEIN, B. A.; HUNNARD, A. E.; CUTLER, A.; BARCELLOS, L F. An application of Random Forests to a genome-wide association dataset: Methodological considerations & new findings. **BMC Genet.**, v.14, p. 11-49, 2010.

HILL, A. V. S. The genomics and genetics of human infectious disease susceptibility. **Annu. Rev. Genomics Hum. Genet.**, v. 2, p. 373-400, 2001.

_____. S. Genetic susceptibility to malaria and other infectious diseases: from the MHC to the whole genome. **Parasitology**, v. 112, p. S75- S84, 1996.

HOFFMANN, K., LINDNER T. H. EasyLINKAGE Plus – Automatic linkage analyses using large-scale SNP data. **Bioinformatics**, v. 21, p. 3565-3567, 2005.

HOSHIDA, Y., FUCHS, B. C., TANABE, K. K. Genomic risk of hepatitis C-related hepatocellular carcinoma. **J. Hepatol.** Disponível em :< <http://www.ncbi.nlm.nih.gov/pubmed/21963516>>. Acesso em 02 Nov. 2011.

JALLOW, M., et al. Genome-wide and fine-resolution association analyses of malaria in West Africa. **Nat. Genet.**, v. 6, p. 657-665, 2009.

JAMES, S. P.; NICOL, W. D.; SHUTE, P. G. A study of induced malignant tertian malaria. **Proc. R. Soc. Med.**, v. 25, p. 1153-1186, 1932.

JASMINE, F. et al. Whole-genome amplification enables accurate genotyping for microarray-based high-density single nucleotide polymorphism array. **Cancer Epidemiol Biomarkes.**, v. 17, n. 12, p. 3499-3508, 2008.

JERONIMO, M. S., et al. Genes at human chromosome 5q31.1 regulate delayed-type hypersensitivity responses with leishmania chagasi infection. **Genes Immun.**, v. 7, p. 539-551, 2007.

JIANG, Y.; WANG, Y.; LIU, T.; LIU, J. Association between HLA class II gene and susceptibility or resistance to chronic hepatitis B. **World J. Gastroenterol.**, v. 9, p. 2221-2225, 2003.

KALLMANN, F. J.; REISNER, D. Twin studies on the significance of genetic factors in tuberculosis. **Am. Rev. Tuberc.**, v. 47, p. 549-574, 1942.

KAWAMATA, C. E. M. **Epidemiologia genética da infecção por *Trypanosoma cruzi***. 2006. 75 f. Dissertação (Mestrado em ciências) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2006.

_____, et al. Genetic epidemiology of the *Trypanosoma cruzi* infection. **J. Hum. Genet.**, v. 15, p. 291-292, 2007.

KLEIN, R. J.; ZEISS, C.; CHEW, E. Y.; TSAI, J. Y.; SACKLER, R. S., HAYNES, C. Complement factor H polymorphism in age-related macular degeneration. **Science**, v. 308, p. 385–389, 2005.

KRIEGER, H. et al. Racial admixture in Northeastern Brazil. **Ann. Hum. Genet.**, Lond., v. 29, p. 113-125, 1965.

_____; FEITOSA, M. F. Genetic epidemiology of infectious disease. **Ciência e Cultura**, v. 51, p. 191-198, 1999.

KRUGLYAK, L. et al. Parametric and non parametric linkage analysis: a unified multipoint approach. **Am. J. Hum. Genet.**, v. 6. p.1347-1363, 1996.

KU, C. S.; LOY, E. Y.;PAWITAN, Y.; CHIA, K. S. The pursuit of genome-wide association studies: where are we now? **J. Hum. Genet.**, v. 55, p. 195-206, 2010.

KUMAR, A. R., et al. Association and mutation analyses of 16p11.2 autism candidate genes. **PloS One.**, v. 4, p. e4582-4590, 2009.

KUROTAKI, N., Identification of novel Schizophrenia loci by homozygosity mapping using DNA microarray analysis. **PloS One.**, v. 6, p. e20589-20590, 2011.

LALOUEL, J. M.; MORTON, N. E. Complex segregation analysis with pointers. **Hum. Hered.**, v. 31, p. 312-321, 1981.

LEHNE, B.; LEWIS, C. M.; SCHLITT, T. From SNPs to genes: Disease Association at the Gene Level. **PLoS ONE**, v.6 p. e20133(10), 2011.

LIGHTART, L. et al. Meta-analysis of genome-wide association for migraine in six population-based European cohorts. **Eur. J. Hum. Genet.**, v.19, p. 901–907, 2011

LIN, T. M. et al. Hepatitis B virus markers in chinese twins. **Anticancer Res.**, v. 9, p. 737-741, 1989.

LIU, H. et al. Repeat-region polymorphisms in the gene for the dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin-related molecule: Effects on HIV-1 Susceptibility. **J. Infect. Dis.**, v. 193, p. 698-702, 2006.

LIU, L., et al. A genome-wide association study with DNA pooling identifies the variant rs11866328 in the GRIN2A gene that affects disease progression of chronic HBV infection. **Viral Immunol.**, v. 24, p. 397-402, 2011.

LIU, Y. et al. Genome-wide interaction-based association analysis identified multiple new susceptibility loci for common diseases. **PLoS Genet.**, v. 7(3), p. e1001338 (16), 2011.

MARTÍNEZ, A., et al. Chromossomal region 16p13; further evidence of increase predisposition to immune diseases. **Ann. Rheum. Dis.**, v. 69, p.309-311, 2010.

MESTRINER, M. A. **Estudo populacional de dois polimorfismos enzimáticos no homem: Esterase-D e Anidrase Carbônica CA II.** 1976. 198 f. Tese (Livre Docência em ciências) - Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto, 1976.

MORO-FURLANI, A. M. **Dois polimorfismos (ES-D e AC-II) no Nordeste do Brasil: estudos populacionais e familiares.** 1977. 77 f. Tese (Mestrado) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1977.

MORTON, N. Sequential tests for the detection of linkage. *American J. Hum. Gen.*, v. 7, p. 277-318, 1955.

MORTON, N. **Outline of genetic epidemiology.** New York: Karger, 1982. 262 p.

NANODROP TECHNOLOGIES. Disponível em: <<http://www.nanodrop.com/Productnd2000overview.aspx>>. Acesso em: 23 fev. 2011.

PENALVA DA SILVA, F. **Dermatoglifos, aspectos demográficos e raciais em uma população rural do Brasil.** 1971. 110 f. Dissertação (Mestrado em ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1971.

_____. **A incompatibilidade ABO no Nordeste brasileiro. Aspectos formais e populacionais.** 1975. X f. Tese (Doutorado em genética) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1975.

_____; KRIEGER, H. The maintenance of ABO polymorfism. *Rev. Bras. Genét.*, v. 4, p. 705-712, 1981.

_____; KRIEGER, H. Haptoglobin segregation and interaction with ABO blood groups. *Rev. Bras. Genét.*, v. 5, p. 157-163, 1982.

ON LINE MENDELIAN INHERITANCE IN MEN. (OMIN). Disponível em: <<http://omim.org/entry/131320>>. Acesso em: 21 set. 2011.

RAMAGOPALAN, S. V.; DYMENT, D. A.; EBERS, G. C. Genetic epidemiology: the use of old and new tools for multiple sclerosis. *Trends Neurosci.*, v. 31 (12), p. 645-652, 2008.

RAMASAWMY, R., et al. *BAT1*, a putative anti-inflammatory gene is associated with chronic chagas cardiomyopathy. *J. Infect. Dis.*, v. 193, p. 1394-1399, 2006a.

_____, et al. The monocyte chemoattractant protein-1 gene polymorphism is associated with cardiomyopathy in human chagas disease. *Clin. Infect. Dis.*, v. 43, p. 305-311, 2006b.

_____, et al. Polymorphism in the gene for Lymphotoxin- α predispose to chronic Chagas cardiomyopathy. *J. Infect. Dis.*, v. 193, p. 1836-1843, 2007a.

_____, et al. Variants in the promotion region of IKBL/NFKBIL1 gene way mark susceptibility to the development of chronic Chagas' cardiomyopathy

among *Trypanosoma cruzi*-infected individuals. **Mol. Immunol.**, v. 45. p. 283-288. 2007b.

_____, et al. Heterozygosity for the S180L variant of *MAL/TTIRAP*, a gene expressing an adaptor protein in the Toll-like receptor pathway, is associated with lower risk of developing chronic chagas cardiomyopathy. **J. Infect. Dis.**, v. 199, p. 1838-1845, 2009.

RAMIREZ, J. D., et al Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic chagasic patients. **PLoS Negl Trop Dis.** v. 4, p. e899-908, 2010.

REY, L. **Parasitologia.** 3^a ed. Guanabara Koogan, 2001. 884 p.

ROSS, K. S., HAITES, N. E., KELLY, K. F. Repeated freezing and thawing of peripheral blood and DNA in suspension: effects on DNA yield and integrity. **J. Med. Genet.**, v. 27, p. 569-570, 1990.

RUSSO, V. E. G. O. **Mistura racial em uma população rural brasileira.** X f. 1972. Dissertação (Mestrado em ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 1972.

SCERRI, T. S., Identification of candidate genes for dyslexia susceptibility on chromosome 18. **PloS One.**, v. 5, p. e13712-13720, 2010.

SIKORA, M., et al. A Targeted Association Study of Immunity Genes and Networks Suggests Novel Associations with Placental Malaria Infection. **PloS One.**, v. 6, p. e24996-25005, 2011.

SPENCER, K. L. et al. Dissection of chromosome 16p12 linkage peak suggests a possible role for CACNG3 variants in age-related macular degeneration susceptibility. **Invest. Ophthalmol. Vis. Sci.**, v. 52, p. 1748-1754, 2011.

SUGIYAMA, M., MIZOKAMI, M. Genome-wide association study and the clinical application to chronic hepatitis C. **Uirusu.**, v. 61, p. 15-24, 2011.

THE WELLCOME TRUST CASE CONTROL CONSORTIUM (WTCCC.). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. **Nature.** v. 447, p. 661–678, 2007.

THYE, T., et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. **Nat. Genet.**, v. 42, p. 739-743. 2010.

TOSH, K. et al. Variants in the SP110 gene are associated with genetic susceptibility to tuberculosis in West Africa. **Proc. Nat. Acad. Sci. U.S.A.**, v. 103, p. 10364-10368, 2006.

TORRES, O. A. et al. Association of the macrophage migration inhibitory factor -173G/C polymorphism with Chagas disease. **Hum. Immunol.**, v. 70, p. 543-546, 2009.

_____ et al. Role of the *IFNG* +874T/A polymorphism in Chagas disease in a Colobian population. **Infect. Genet. Evol.**, v. 10, p. 682-685, 2010.

WALTERS, R. G., A novel highly-penetrant form of obesity due to microdeletions on chromosome 16p11.2. **Nature.**, v. 463, p 671-675, 2010.

WILLIAMS-BLANGERO, S.; VANDEBERG J. L.; BLANGERO J.; CORRÊA-OLIVEIRA R. Genetic epidemiology of *Trypanosoma cruzi* infection and chagas' disease. **Frontiers Biosci.**, v. 8, p. e337-e345, 2003.

WORLD HEALTH ORGANIZATION, (WHO). Disponível em: <<http://www.who.int/mediacentre/factsheets/fs340/en/index.html>>. Acesso em: 12 jul 2010.

APÊNDICE

Anexo 1 – Análise de regressão múltipla escalonada

Variáveis	Coefficientes não padronizados	Erro Padrão
intercepto	0,0031450	0,007
IDADE	0,0067080	0,001
(IDADE) ²	-0,0000894	0,000
TIPO DE HABITAÇÃO	-0,0149700	0,004

Anexo 2 – Análise de correlação

Tipo de par	N	r	Significância
PAI X MÃE	633	0,153	P < 0,01
PAI X FILHO	2648	0,186	P < 0,01
PAI X FILHO médio	695	0,203	P < 0,01
MÃE X FILHO	2697	0,166	P < 0,01
MÃE X FILHO médio	726	0,203	P < 0,01
PAIS média X FILHOS	2413	0,217	P < 0,01
PAIS média X FILHO médio	627	0,261	P < 0,01
FILHOS (intraclasse)	3209	0,224	P < 0,01

Anexo 3 – Análises de componentes de variância

Ambiente Comum	0,523
Dominante	0,101
Aditivo	
esperado ao acaso	0,353
por casamento preferencial	0,022

Anexo 4 – Análise de segregação complexa

Modelo	d	t	q	H	r_1	r_2	r_3	-2 ln L	χ^2	P	teste	PE	AIC
1.Misto*	1,000	2,096	0,605	0,421	[1]	[0,5]	[0]	672,40				4	680,397
2.Aleatório	[0]	[0]	[0]	[0]	-	-	-	838,63	166,238	0,000	2 vs. 1	0	838,635
3.Sem gene principal **	[0]	[0]	[0]	0,995	[1]	[0,5]	[0]	696,37	23,974	0,000	3 vs. 1	1	698,372
4.Sem comp. multifatorial	0,097	3,250	0,296	[0]	[1]	[0,5]	[0]	712,27	39,869	0,000	4 vs. 1	3	718,266
5.Recessivo (d=0)	[0]	2,389	0,383	0,279	[1]	[0,5]	[0]	682,78	10,381	0,001	5 vs. 1	3	688,778
5'.Recessivo (d=0, H=0)	[0]	2,999	0,309	[0]	[1]	[0,5]	[0]	712,75	29,975	0,000	5 vs. 5'	2	716,753
6.Codominante (d=0,5)	[0,5]	2,782	0,454	0,019	[1]	[0,5]	[0]	674,09	1,688	0,194	6 vs. 1	3	680,085
7. Dominante (d=1)	[1]	2,096	0,605	0,421	[1]	[0,5]	[0]	672,40	0,000	1,000	7 vs. 1	3	678,397
8.ts livre (d=1)	[1]	1,002	0,319	0,750	1,000	0,720	-0,553	681,44	9,046	0,029	8 vs. 7	6	693,443
9.ts iguais (d=1) **	[1]	1,000	1,000	0,800	0,900	0,900	0,900	714,84	42,446	0,000	9 vs. 7	6	726,843

* Valor de "d" foi fixado no limite superior.** Modelo estimado por "-2 ln L mínimo". Parâmetros entre colchetes foram fixados no valor indicado.

d = grau de dominância; t = deslocamento; q = frequência alélica; H = herdabilidade multifatorial;

r_s = parâmetros de probabilidades de transmissão (#8 o modelo com r_s livres, enquanto no modelo #9 os r_s são iguais);

-2 ln L = menos duas vezes o logaritmo neperiano do likelihood (χ^2). χ^2 = qui-quadrado entre os modelos; PE = parâmetros estimados.

Anexo 5 – Protocolo de genotipagem Affymetrix

1. Preparação do DNA

LOCAL: LABORATÓRIO

- Quantificar o DNA no NanoDrop,
- Aliquotar 5 µl por amostra equivalentes a 250 ng, 50 ng/µl com tampão de *EDTA TE* (10 mM Tris HCl, 0,1 mM EDTA, pH 8,0).
- Centrifugação rápida e guardar a -20 °C ou prosseguir para o próximo passo.

2. Digestão

LOCAL: SALA LIMPA

Preparação na capela, matérias necessários:

- *Cooler* e gelo,
- Nomear tubo *ependorf* de 1,5 ml para o mix de digestão e tubos de 0,2 para cada amostra,

- Adicionar amostras de controle positivo (*Reference Genomic DNA 103*) e controle negativo (H₂O),
- Colocar a água Molecular Grade no gelo,
- Colocar amostras no gelo para descongelar, *vortex* e *spin*,
- Ligar e pausar o termociclador no programa 500K-DIG, (ver Anexo 5)
- Centrifugação rápida na enzima *Nsp I* ou *Sty I* e colocar no cooler,
- Colocar o *NE Buffer 2 10 X (Nsp I)* e o *BSA (100 X; 10 mg/ml)* no gelo para descongelar.

Preparação da solução de digestão na capela:

Água Molecular Grade	11,55 µl
----------------------	----------

<i>NE Buffer 10X (Nsp I ou Sty I)</i>	2 µl
---------------------------------------	------

<i>BSA (100X; 10 mg/ml)</i>	0,2 µl
-----------------------------	--------

<i>Nsp I (10 U/µl)</i>	1 µl
------------------------	------

Total	14,75 µl
-------	----------

- Manter o mix de digestão no gelo.
- Distribuir o mix de digestão nas amostras na bancada:
- Adicionar 14,75 µl do mix de digestão aos 5 µl de amostra, total 19,75 µl
- Homogeneizar e centrifugar rapidamente.
- Colocar no termociclador e continuar o programa pausado 500K-DIG. (~2,5 h)
- Centrifugar rapidamente e guardar a -20 °C ou prosseguir para o próximo passo.

3. Ligação

LOCAL: SALA LIMPA

Preparar na capela:

- Cooler e gelo,
- Nomear tubo *eppendorf* de 1,5 ml para o mix de ligação,
- Colocar amostras no gelo para descongelar, homogeneizar e centrifugar rapidamente,
- Tampão *T4 DNA Ligase* 10 X (usar alíquota e não descongelar cada alíquota mais do que 2 vezes) e Adaptador *Nsp I*, descongelar no gelo, homogeneizar e centrifugar rapidamente,
- Centrifugar rapidamente o *T4 DNA Ligase* (400 U/μl; NEB) e colocar no cooler,
- Ligar e pausar o termociclador no programa 500K-LIG.(ver Anexo 5)

Preparação da solução de ligação na capela:

<i>Adaptor Nsp I</i> (50 μ M)	0,75 μ l
Tampão <i>T4 DNA Ligase</i> (10 X)	2,5 μ l
<i>T4 DNA Ligase</i> (400 U/ μ l)	2 μ l
<hr/>	
Total	5,25 μ l

- Manter a solução de ligação no gelo,
- Homogeneizar as amostras.

Distribuir a solução de ligação na bancada:

- Adicionar 5,25 μ l da solução de ligação aos 19,75 μ l das amostras de DNA digerido, total 25 μ l,
- Homogeneizar e centrifugar rapidamente.
- Colocar no termociclador e continuar o programa pausado 500K-LIG. (~3,5 h),
- Centrifugar rapidamente e guardar a -20 °C ou prosseguir para o próximo passo.

4. Diluição

LOCAL: SALA LIMPA

Fazer a diluição na bancada:

- Colocar a água Molecular Grade no gelo,
- Colocar amostras no gelo para descongelar, homogeneizar e centrifugar rapidamente,
- Adicionar 75 µl de água Molecular Grade nas amostras, total 100 µl,
- Homogeneizar e centrifugar rapidamente.

Spin e guardar a -20 °C ou prosseguir para o próximo passo.

5. PCR

LOCAL: LABORATÓRIO

- Ligar e pausar o termociclador no programa 500K-PCR, (ver anexo 5)
- Preparar um gel de agarose 2%.

LOCAL: SALA LIMPA

Preparação na capela:

- *Cooler* e gelo,
- Colocar a água Molecular Grade e *GC-Melt* no gelo,
- Nomear tubo *falcon* de 15 ml ou 50 ml para o mix da PCR e 3 tubos de 0,2 para cada amostra,
- Tampão *TITANIUM Taq PCR*, dNTPs e *PCR Primer 002* para descongelar no gelo, homogeneizar e centrifugar rapidamente,
- Colocar amostras no gelo para descongelar, homogeneizar e centrifugar rapidamente,
- Manter a *TITANIUM Taq DNA Polymerase* no congelador e só retirar na hora do uso.

- Preparação da solução de PCR:

Água Molecular Grade	39,5 µl
Tampão <i>TITANIUM Taq PCR</i> (10 X)	10 µl
<i>GC-Melt</i> (5 M)	20 µl
<i>dNTP</i> (2,5 mM cada)	14 µl
<i>PCR Primer 002</i> (100 µM)	4,5 µl
<i>TITANIUM Taq DNA Polymerase</i> (50 X)	2 µl
<hr/>	
Total	90 µl

- Manter a solução de PCR no gelo.
- Distribuir a solução de PCR nas amostras no gelo na bancada:
- Fazer triplicatas de 10 µl de cada amostra, e colocar o resto no congelador,
- Adicionar 90 µl da solução de digestão aos 10 µl de amostra, total 100 µl,
- Homogeneizar e centrifugar rapidamente.
- Colocar no termociclador e continuar o programa pausado 500K-PCR (~2,5 h),

- Centrifugar rapidamente,
- Guardar a -20 °C ou prosseguir para o próximo passo.

6. Gel de Verificação 1

LOCAL: LABORATÓRIO

- Correr 3 µl de cada amostra com 1 µl de corante no gel de agarose 2% por mais ou menos 1 h 15 min. a aproximadamente 120 V. Reservar um pocinho para 5 µl de peso molecular,
- Verificar a distribuição do produto da PCR no gel (250 bp até 1100 bp).

7. Purificação

LOCAL: LABORATÓRIO

Preparação das amostras:

- Separar as amostras para descongelar, homogeneizar e centrifugar rapidamente,
- Descartar o controle negativo,
- Diluir o EDTA para 0,1 M (estoque 5 M, pH 8,0),
- Acrescentar 8 µl de EDTA 0,1M em cada amostra, *vortex* e *spin*,
- Juntar as 3 reações de PCR em cada pocinho da placa de purificação,
- Colocar um lenço de papel sobre a placa de purificação para evitar sujeiras,
- Fazer um mapa da placa de purificação.

Preparação da bomba de vácuo:

- Montar aparelhos para ligar na bomba de vácuo,
- Colocar a placa com as amostras no aparelho da bomba de vácuo,
- Ligar a bomba de vácuo com a válvula aberta e fechar devagar até atingir 600 mbar,

- Lavar cada pocinho 3 vezes com 50 μ l de água Molecular Grade,
- Esperar secar completamente (~1 h 30 min.),
- Colocar um lenço de papel sob a placa de purificação para secar completamente,

Não demorar mais do que 90 minutos para começar o próximo passo.

8. Eluição

LOCAL: LABORATÓRIO

- Adicionar 45 µl de tampão *RB* em cada pocinho da placa de purificação,
- Selar bem a placa de purificação com o adesivo de alumínio,
- Colocar a placa de purificação no agitador a aproximadamente 1000 rpm por 10 minutos,
- Nomear novos tubos,
- Transferir os 45 µl de cada amostra da placa de purificação para os novos tubos,
- Centrifugar rapidamente e guardar a -20 °C ou prosseguir para o próximo passo.

9. Quantificação

LOCAL: LABORATÓRIO

Preparação das amostras para quantificação:

- Separar as amostras para descongelar, homogeneizar e centrifugar rapidamente,
- Quantificar o DNA no NanoDrop.
- Guardar a -20 °C ou prosseguir para o próximo passo.

10. Normalização

LOCAL: LABORATÓRIO

- Normalizar cada amostra para 2 µg/µl com tampão *RB*,
- Separar 45 µl (com 90 µg de DNA) de cada amostra,
- Centrifugar rapidamente e guardar a -20 °C ou prosseguir para o próximo passo.

11. Fragmentação

LOCAL: LABORATÓRIO

Preparação das amostras para Fragmentação:

- Colocar a água Molecular Grade no gelo,
- Separar as amostras para descongelar, homogeneizar e centrifugar rapidamente,
- Descongelar o tampão de Fragmentação (10 X) no gelo, homogeneizar e centrifugar rapidamente,
- Preparar um gel de agarose 4%,
- Ligar e pausar (por pelo menos 10 min.) o termociclador no programa 500K-FRG, (ver Anexo 5)
- Separar e nomear um tubo tipo *ependorf* para a solução de fragmentação,
- Adicionar 5 µl do tampão de Fragmentação (volume final 50 µl).

Preparação da diluição do reagente de Fragmentação:

- Retire o reagente de Fragmentação do congelador somente no momento de uso, homogeneizar e centrifugar rapidamente,

- Dilua o reagente de Fragmentação para 0,05 U/μl usando o tampão de Fragmentação e água Molecular Grade, de acordo com a fórmula a seguir;

$$X = \frac{0,05 \text{ U/}\mu\text{l} \times Y \mu\text{l}}{Z \text{ U/}\mu\text{l}}$$

Onde:

X = volume inicial (μl) do *Fragment Reagent*,

0,05 U/μl = concentração final do *Fragment Reagent* na diluição,

Y = volume final do *Fragment Reagent* (5 μl em cada amostra),

Z = concentração inicial (U/μl) do *Fragment Reagent* em estoque.

Tabela da diluição:

H ₂ O Molecular Grade	$Y - (0,1 Y + X)$
10 X Fragmention Buffer	0,1 Y
Fragment Reagent	X
<hr/>	
Total	Y

- Acionar 5 µl do reagente de Fragmentação diluído a cada amostra (volume final 55 µl), o mais rápido possível, uma vez que o mesmo perde a atividade mesmo no gelo,
- Homogeneizar e centrifugar rapidamente a 4 °C a 2000 rpm por 30 segundos,
- Colocar no termociclador e continuar o programa pausado 500K–FRG (~0,75 h).
- Centrifugar rapidamente e guardar a -20 °C ou prosseguir para o próximo passo.

12. Gel de Verificação 2

LOCAL: LABORATÓRIO

- Correr 4 µl de cada amostra com 4 µl de corante no gel de agarose 4% por mais ou menos 45 minutos a aproximadamente 110 V. Reservar um pocinho para 5 µl de peso molecular,
- Verificar a distribuição do produto da fragmentação no gel (<180bp).

13. Marcação

LOCAL: LABORATÓRIO

- Descongelar o tampão *TdT* 5 X e o *GeneChip® DNA Labeling Reagent* no gelo, homogeneizar e centrifugar rapidamente,
- Deixar a enzima *TdT* no congelador até o momento do uso,
- Nomear os tubos das amostras e um tubo grande para a solução e deixar no *cooler*,
- Ligar e pausar (por 10 min.) o termociclador no programa 500K-LAB, (ver Anexo 5)

Preparação da solução de marcação:

Tampão <i>TdT</i> 5 X	14 µl
<i>GeneChip® DNA Labeling Reagent</i>	2 µl
Enzima <i>TdT</i> (30 U/µl)	3,5 µl
<hr/>	
Total	19,5 µl

- Manter a solução de marcação e as amostras no gelo,
- Adicionar 19,5 μ l do mix de marcação aos 50,5 μ l (51 μ l) de amostra, o volume total é 70 μ l,
- Homogeneizar e centrifugar rapidamente,
- Colocar no termociclador e continuar o programa pausado 500K-LAB,
- Centrifugar rapidamente e guardar a -20 °C ou prosseguir para o próximo passo.

14. Hibridação

LOCAL: LABORATÓRIO

Preparação do 12 X MES (1,25 M MES, 0,89 M [Na⁺]) de hibridação (1 l):

- 70,4 g *MES Hydrate*,
- 193,3 g *MES Sodium Salt*,
- 800 ml de água *AccuGENE®*,
- Completar para 1 l (pH deve estar entre 6,5 e 6,7),
- Montar o kitassato na bomba de vácuo com filtro de 0,2 µm,
- Ligar a bomba de vácuo com a válvula aberta e fechar devagar até atingir 600 mbar,
- Guardar na geladeira e proteger da luz, descartar se ficar amarelado.

Preparação do mix de hibridação:

<i>MES</i> (12 X; 1,22 M)	12 µl
<i>DMSO</i> (100%)	13 µl
<i>Denhardt's Solution</i> (50X)	13 µl
<i>EDTA</i> (0,5 M)	3 µl
<i>HSDNA</i> (10 mg/ml)	3 µl
<i>OCR</i> , 0100	2 µl
<i>Human Cot-1 DNA</i> ® (1 mg/ml)	3 µl
<i>Tween-20</i> (3%)	1 µl
<i>TMACL</i> (5 M)	140 µl
<hr/>	
Total	190 µl

- Centrifugar rapidamente guardar a -20 °C ou prosseguir para o próximo passo.
- Tirar os chips da geladeira e deixar a temperatura ambiente.
- Nomear novos tubos de 1,5 ml.
- Adicionar 190 µl da solução de hibridação aos 70 µl da amostra.

- Colocar a solução de hibridação e o DNA marcado a 96 °C por 10 minutos.
- Colocar no gelo por 10 segundos.
- Centrifugar rapidamente.
- Colocar as amostras a 49 °C por pelo menos 1 minuto.
- Nomear os chips e perfurar o septo superior com uma ponteira P200.
- Injetar 200 µl da amostra pelo septo de baixo.
- Colocar os chips no forno a 49 °C por 16 a 18 horas a 60 rpm.
- Tirar os chips do forno de hibridação e retirar os 200 µl do chip.

Preparação de 100 ml do tampão 1 X *Array Holding* (100 mM MES, 1 M [Na⁺], 0,01% Tween 20):

- 8,3 ml de tampão 12 X MES *Stock*,
- 18,5 ml de 5 M NaCl,
- 0,1 ml de 10% Tween 20,
- 73,1 ml de água Miliq,
- Guardar na geladeira.
- Encher o chip com 270 µl de tampão *Array Holding*.
- Prosseguir para o próximo passo imediatamente ou manter os chips na geladeira por 3

15. Lavagem

LOCAL: LABORATÓRIO

Preparação de 1 l do reagente *Wash A* (6 X SSPE, 0,01% Tween 20):

- 300 ml de 20 X SSPE,
- 1,0 ml de 10% Tween 20,
- 699 ml de água Miliq,
- Montar o kitassato na bomba de vácuo com filtro de 0,2 μ m,
- Ligar a bomba de vácuo com a válvula aberta e fechar devagar até atingir 600 mbar,
- Manter a temperatura ambiente,
- Posicionar o frasco no local indicado ao lado da *Fluidics Station* e conectar a mangueira.

Preparação de 1 l do reagente *Wash B* (6 X SSPE, 0,01% Tween 20):

- 30 ml de 20X SSPE,
- 1,0 ml de 10% Tween 20,
- 969 ml de água Miliq,

- Montar o kitassato na bomba de vácuo com filtro de 0,2 μm ,
- Ligar a bomba de vácuo com a válvula aberta e fechar devagar até atingir 600 mbar,
- O pH deve ser igual a 8,
- Manter a temperatura ambiente,
- Posicionar o frasco no local indicado ao lado da *GeneChip Fluidics Station 450* e conectar a mangueira.

Preparação do *Anti-Streptavidin Antibody* 0,5 mg/ml:

- Ressuspender 0,5 mg do anticorpo *Anti-Streptavidin* em 1 ml de água Miliq,
- Guardar a 4 °C.

Preparação de 1 l do 12 X MES (1,25 M MES, 0,8 9M [Na⁺]):

- 70,4 g MES *Hydrate*,
- 193,3 g MES *Sodium Salt*,
- 800 ml de água *AccuGENE®*,
- Completar para 1 l (pH deve estar entre 6,5 e 6,7),
- Montar o kitassato na bomba de vácuo com filtro de 0,2 μm ,

- Ligar a bomba de vácuo com a válvula aberta e fechar devagar até atingir 600 mbar,
- Guardar na geladeira e proteger da luz, descartar se ficar amarelado.

Para registrar as amostras a serem analisadas:

- Clicar no ícone *Affymetrix Laucher* localizado na área de trabalho do computador conectado a plataforma de *arrays*,
- Clicar no ícone *AGCC Portal*,
- Clicar em “Samples – Registration”,
- Em “Detailed Sample Registration”, ir para “Select Project” e selecionar o projeto (Ex. Hospedaria),
- Em “Sample File Name”, colocar o número da amostra (Ex. Hosp01),
- Em “Array”, colocar no número do código de barras do chip,
- No final da página clicar em “Salvar”.
-

Preparação do tampão *Staining*:

Água miliq	800,04 µl
------------	-----------

SSPE (20 X)	360 µl
-------------	--------

Tween 20 (3%)	3,96 µl
---------------	---------

Denhardt's (50 X)	24 µl
-------------------	-------

Total	1188 µl
-------	---------

Preparação da solução *SAPE*:

Tampão <i>Stain</i>	594 µl
---------------------	--------

1 mg/ml <i>Streptavidin Phycoerythrin</i> (SAPE)	6,0 µl
--	--------

Total	600 µl
-------	--------

- Centrifugar rapidamente.

Preparação da *Antibody Stain Solution*:

Tampão <i>Stain</i>	594 μ l
Anticorpo 0,5mg/mL <i>biotinylated</i>	6,0 μ l

Total	600 μ l
-------	-------------

- Nomear tubos tipo *ependorf* de 1,5 ml de 1 a 3 para cada módulo da estação de lavagem.
- No tubo número 1, adicione 600 μ l da solução *SAPE*.
- No tubo número 2, adicione 600 μ l de *Antibody Stain Solution*.
- No tubo número 3, adicione 820 μ l de tampão *Array Holding*.
- Colocar os tubos nos locais indicados em cada módulo.
- Para iniciar a lavagem dos chips:
- Clicar no ícone *Affymetrix Lancher* localizado na área de trabalho do computador conectado a plataforma de *arrays*,
- Clicar no ícone *AGCC Fluidics Station* e vá para a aba “Master”,
- No quadro “Step 1” selecionar “Mapping 250K_Nsp.Mapping” em “Probe Array Type”,

- No quadro “Step 2” selecionar “Mapping500K1v2_450” em “Protocol”,
- Selecionar a caixa “Check/Uncheck All Stations And Modules” ao lado do quadro “Step 2”,
- No quadro “Step 3” clicar em “Copy To Selected Modules”,
- Na aba “Station 1”, selecionar a amostra em “Sample File” em cada módulo,
- Clicar em “Run” em cada módulo (30 a 45 minutos),
- Retirar o chip da *GeneChip Fluidics Station 450* e verificar se não existe nenhuma bolha de ar (caso exista, coloque o novamente o chip no mesmo módulo e aguarde o *GeneChip Fluidics Station 450* preenchê-lo corretamente ou complete manualmente com 270 µl de *Array Holding Buffer*).
- Tampar os septos com um adesivo para evitar vazamentos.
- Escanear imediatamente ou guardar na geladeira embrulhado em papel alumínio por no máximo 24 horas).

16. Escaneamento

LOCAL: LABORATÓRIO

- Retirar da geladeira o chip a ser escaneado 20 minutos antes do início do escaneamento.
- Clicar no ícone *Affymetrix Lancher* localizado na área de trabalho do computador conectado a plataforma de microarranjos.
- Abrir o programa *AGCC Scan Control* e adicionar o número do indivíduo.
- Clicar no botão *Start*.

17. Verificação do chip escaneado

LOCAL: LABORATÓRIO

- Clicar no ícone *Affymetrix Viewer* localizado na área de trabalho do computador conectado a plataforma de *arrays*.
- Verificar o controle positivo nos quatro cantos de cada chip e a escrita “GeneChip” nos arranjos, caso não estejam luminescentes a amostra deve ser refeita.

Genes em indivíduos e famílias

4

Ricardo G. M. Ferreira
Carlos E. M. Kawamata
Henrique Kriger

INTRODUÇÃO

Atualmente, na época chamada de “pós-genômica”, o enfoque dos estudos em genética humana, de uma maneira geral, tem-se voltado para os fenômenos de causa, e efeito de genes conhecidos na produção de certas características nosológicas particulares e fenotípicas. Por outro lado, muitas características humanas não possuem um mecanismo etimológico simples e são causadas pelo efeito simultâneo de vários fatores biológicos e do ambiente. Essas características são chamadas “complexas”, pois não existe um mecanismo (genético ou ambiental) simples que explique o desenvolvimento da característica em estudo. Como exemplo de característica complexa podem ser citadas as doenças cardiovasculares, em que um número grande de fatores afeta o fenótipo (tipo de alimentação, exercícios, hábito de fumar, excesso de peso e um componente genético importante, porém de natureza ainda não esclarecido).

No presente capítulo, serão tratados apenas aspectos importantes nos métodos e nas abordagens específicas do estudo de características complexas. Existem bons livros dedicados ao tratamento de dados sobre características monogênicas e que o leitor buscando informações mais complexas deve-se

valer (conferir Beiguelman, “A interpretação genética da variabilidade humana” na página <http://www.desvirtual.com/bbeiguel/ebook03.htm>).

Com o fabuloso desenvolvimento da biologia molecular e conseqüente desenvolvimento de métodos e técnicas, equipamentos e treinamento de profissionais competentes, no começo do presente milênio, estabeleceu-se com quase absoluta precisão, o genoma humano com todos os genes que compõem esse patrimônio hereditário e suas seqüências de nucleotídeos (ver Capítulo 2 – Genes e cromossomos).

Porém, deve-se salientar que o gene não possui apenas essa propriedade (i.e. uma determinada seqüência), mas também outras que são de extrema importância para os estudos de características complexas e, na maioria, bastante simples e conhecidas há muitas décadas. Essas propriedades, apesar de poderem ser enunciadas e entendidas facilmente, possuem especificidades matemáticas bastante claras e vêm, com o constante aprimoramento de métodos, colaborando para o conhecimento das contribuições relativas de várias causas em características normais e patológicas da espécie.

A primeira propriedade do gene que será tratada aqui é a *transmissão* da característica de

uma geração para a geração subsequente. Como se sabe da genética clássica, existem apenas três situações distintas de um gene passar ou não passar de um genitor para sua prole.

- O referido genitor é homozigoto para o gene. Nesse caso, toda a prole receberá, o referido gene. Pode-se simbolizar essa transmissão em termos probabilísticos da seguinte forma: $P(AA \rightarrow A) = 1$, ou seja, a probabilidade de um indivíduo AA transmitir o gene A é um, desde que tenha descendência.
- O genitor é heterozigoto para o gene A (genótipo AA'). Nesse caso, a probabilidade de transmissão do gene A é 0,5. Simbolicamente: $P(AA' \rightarrow A) = 0,5$.
- O genitor não possui o gene A. Obviamente podemos dizer que $P(A'A' \rightarrow A) = 0$.

Nos trabalhos que tratam de analisar o comportamento dos genes em famílias, essas três probabilidades possuem geralmente, as seguintes notações $P(AA \rightarrow A) = \tau_{AA}$, $P(AA' \rightarrow A) = \tau_{Aa}$ e $P(A'A' \rightarrow A) = \tau_{aa}$.

A segunda propriedade que será abordada é o *efeito*. Para se visualizar a maneira pela qual é quantificado o efeito, utiliza-se uma abstração relativamente simples: imagina-se que o indivíduo homozigoto AA produz em média um efeito t sobre o fenótipo em estudo, ao passo que o indivíduo A'A' não produz nenhum efeito visível no fenótipo ($t = 0$). O efeito produzido pelo heterozigoto (AA') é igual a dt , em que d é o que se denomina de *dominância*. Dessa forma o gene A humano é dominante quando $d = 1$, e portanto, o efeito do gene no heterozigoto (AA') é igual ao efeito no homozigoto (AA). Quando o valor de d é zero, diz-se que o gene A é recessivo, uma vez que os fenótipos de A'A' e AA' são iguais. Existe dominância parcial quando $0 < d < 1$. Quando $d = 0,5$, há co-dominância.

Uma outra propriedade importante do gene é sua *frequência* populacional (q). Essa

propriedade é definida por um parâmetro clássico estabelecido no começo do século XX por Hardy & Weinberg e que costuma ser estimada por simples contagem gênica ou por intermédio de simplificações algébricas quando existe dominância.

Esses parâmetros (τ , t , d e q) são geralmente estimados simultaneamente no que hoje se denomina análise de segregação complexa, conjuntamente com outro parâmetro (H) que é a *herdabilidade* no sentido tradicional que mede a combinação total do componente multifatorial que inclui o conjunto de poligenes que atuam sobre a característica em estudo. Finalmente, mais uma propriedade do gene de importância capital para os estudos de epidemiologia genética é a *recombinação*, que é medida pela taxa de recombinação (θ) e definida como a relação entre o gene em estudo e um determinado marcador. Essa última propriedade será tratada com mais pormenores a seguir.

LIGAÇÃO GÊNICA

Breve perspectiva histórica

Após a publicação dos trabalhos de Charles Darwin em meados do século XIX, pesquisadores de diversas áreas das Ciências Naturais, especialmente aqueles interessados em melhoramento vegetal, demonstraram grande interesse em uma abordagem prática da Teoria da Evolução. O trabalho de Mendel, publicado em fevereiro de 1866, intitulado *Versuche über Pflanzen-Hybriden* – Experimentos na Hibridação de Plantas, foi recebido com entusiasmo pela comunidade científica, depois de traduzido para o inglês por Willian Bateson (1901). O artigo original de Mendel, assim como a tradução de Bateson e outros textos muito interessantes sobre o início da Genética Moderna podem ser encontrados na página <http://www.mendelweb.org/>.

Logo depois da sua redescoberta, as Leis de Mendel foram testadas por diversos pesquisa-

dores, e alguns deles começaram a encontrar proporções entre dois fenótipos dominantes que segregam independentemente quando analisados separadamente, diferentes das esperadas na geração F_2 de acordo com a Segunda Lei de Mendel (9:3:3:1), quando analisados de forma conjunta. Como exemplo, tem-se o estudo de Bateson e Punnett (1911) que utilizaram ervilhas. Apesar das observações, os dois pesquisadores não forneceram uma explicação para o fenômeno, chamado no momento de *coupling* e *repulsion*. Estudando drosófilas, Thomas Hunt Morgan publicou uma série de trabalhos sobre os mecanismos hereditários, entre eles o livro *Mechanism of Mendelian Heredity* (1915), no qual propõe que os genes estão organizados linearmente em cromossomos, e desse modo, fornecendo uma explicação sobre as divergências das frequências fenotípicas observadas quando comparadas àquelas esperadas pelas Leis de Mendel.

Considerando os trabalhos de Morgan e sua escola, ficou claro que nem todos os genes segregam de forma independente. Genes que estão em um mesmo cromossomo estão ligados uns aos outros. Uma visão probabilística da ligação entre os genes deu origem ao conceito de *ligação gênica*.

Em humanos, um exemplo clássico de ligação entre os genes é o que forma o sistema principal de histocompatibilidade (MHC).

Conceito atual de ligação gênica

Do ponto de vista celular, os genes de um mesmo cromossomo estão ligados entre si. Em cada célula somática, um indivíduo possui duas cópias homólogas ($2n$) de cada cromossomo, uma proveniente de seu pai e outra de sua mãe. Durante a meiose, a divisão celular que leva à formação dos gametas, cada um dos cromossomos homólogos é duplicado, formando quatro cromátides (duas paternas e duas maternas). Posteriormente, essas cromátides são pareadas; nesse momento, frequente-

mente as cromátides pareadas não se separam completamente em alguns pontos, formando os quiasmas. Nos pontos onde ocorrem os quiasmas, segmentos do cromossomo de origem paterna são trocados por maternos, formando um cromossomo com uma nova combinação de alelos, chamado de recombinante. As recombinações ocorrem nos pontos onde ocorrem os quiasmas. Depois das duas divisões da meiose, os gametas terminam com uma cópia de cada cromossomo (n), porém essas cópias normalmente são diferentes daquelas encontradas em células somáticas, devido aos cromossomos recombinantes.

Uma vez que a formação de um quiasma inibe a formação de outro muito próximo, sua formação ocorre em posições semi-aleatórias (Watson et al., 1987). Isso torna possível a construção de mapas genéticos. Genes próximos uns aos outros têm, portanto, uma menor chance de recombinação entre si. Por definição, uma unidade de distância em um mapa genético é igual a 1% de recombinação entre dois genes. Essas unidades são comumente chamadas de centimorgan (cM), em homenagem ao pesquisador Morgan. Nos grandes cromossomos humanos, é frequente a ocorrência de mais de uma recombinação. No cromossomo 1, por exemplo, ocorrem em média quatro recombinações por mitose, o que possibilita uma distância de mais de 100 cM entre dois genes que se encontram nos extremos de um cromossomo.

Existem diversos métodos para detecção de ligação entre dois ou mais genes. Classicamente, o parâmetro τ , que pode variar de 0 a 0,5 cM é calculado medindo-se o número de recombinações entre dois genes na formação dos gametas. O parâmetro θ assume o valor máximo de 0,5 cM porque uma célula dá origem a quatro gametas, dos quais apenas dois podem ser recombinantes.

A ligação é, portanto, uma propriedade do gene, assim como a sequência, que se refere às relações entre os genes. O mapa de ligação traz os genes ordenados com suas respec-

tivas distâncias x baseadas na recombinação. A distância entre genes sucessivos é aditiva, ou seja, dado quatro genes A, B, C e D localizados nessa ordem em um único cromossomo, a distância x entre A e D é $\approx (x_{AB} + x_{BC} + x_{CD})$. A equação mais simples para a construção de mapas de ligação é $x = \tau$, onde x é a distância entre os genes. Essa equação, conhecida como *função de mapeamento de Morgan* (Morgan, 1928), tende a superestimar a distância entre genes mais distantes.

Na tentativa de resolver as superestimativas de distância decorrentes da equação proposta por Morgan e levar em conta os fenômenos que alteram a linearidade de x , foram sugeridas outras equações. Dentre as mais usadas, pode-se citar a de Haldane (1919), binomial (Karlin, 1984) e a de Kosambi (Kosambi, 1944). Na página <http://www.ncbi.nlm.nih.gov/projects/genome/guide/human/>, pode-se encontrar mapas de ligação de todos os cromossomos humanos, além dos mapas físicos e muitas outras informações sobre o genoma humano.

Nos casos de genes humanos, os estudos de ligação têm uma complicação adicional, pois não é possível realizar cruzamentos entre indivíduos com características “interessantes”, além de haver maior intervalo entre gerações. Para isso foram criados diversos métodos de estudo baseados principalmente na obtenção de dados familiares. Com o uso de dados de indivíduos relacionados é possível, em condições favoráveis, inferir a ocorrência de recombinações e até determinar a localização de um gene relacionado ao fenótipo em estudo delimitado por genes marcadores. Os marcadores são genes de localização conhecida, nos quais o genótipo pode ser inferido com razoável precisão a partir dos fenótipos.

Antes dos avanços trazidos pela biologia molecular nos últimos 25 anos, os pesquisadores contavam com cerca de 60 marcadores genéticos clássicos, por meio dos quais realizavam a busca por novos genes. Nesse cenário,

pesquisadores como o professor McKusick, que iniciou em 1966 um dos maiores bancos de dados de genes humanos, o livro *Mendelian Inheritance in Man* (MIM) foi publicado a primeira vez em 1966 e atualmente conta com uma versão *on-line*, o OMIM (*On-line Mendelian Inheritance in Man*). A Tabela 4.1 retirada do portal do OMIM (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>), mostra as estatísticas atuais sobre os genes catalogados nessa página.

Entre os métodos de análise de ligação propostos, um dos pioneiros e que vêm sendo amplamente utilizado é o método do *lod score* proposto por Newton E. Morton (1955), um dos pesquisadores teóricos mais importantes na proposição de métodos de análise genética em humanos, com amplo embasamento estatístico.

Atualmente, com a disponibilidade de diversos painéis de marcadores de STRs ou SNPs que cobrem desde regiões específicas do genoma até o genoma inteiro, a informação genética dos indivíduos e das populações ficou muito mais acessível.

A possibilidade de obtenção de dados criou uma demanda por novos métodos de análise de ligação. O principal problema a ser acessado decorre do grande número de testes aplicados sobre uma mesma amostra. Imaginado um estudo do tipo de escaneamento genômico (*genome scan*) com cerca de 400 marcadores de STRs e uma análise de ligação dois pontos, encontra-se pelo menos 400 testes. Se o nível de significância α de 0,05 for adotado, esperaria-se cerca de 20 resultados positivos espúrios, um erro do tipo I muito alto. Por outro lado, a adoção de uma correção simples como a de Bonferroni, que basicamente é a divisão de α pelo número de testes realizados para obtenção de um α_c , que nesse exemplo seria igual a 0,000125, tornaria praticamente impossível a obtenção de um *lod score* com essa significância, tornando o erro do tipo II muito alto.

Tabela 4.1 Estatísticas do OMIM em 21 de novembro de 2006 – números de registros

Genes	Autossômicos	Ligados ao X	Ligados ao Y	Mitocondriais	Total
*Genes com sequência conhecida	10542	488	48	37	11115
+Genes com sequência e fenótipos conhecidos	354	33	0	0	387
#Descrição de fenótipo, bases moleculares conhecidas	1821	161	2	26	2010
%Fenótipo mendeliano, bases moleculares desconhecidas	1415	137	4	0	1556
Outros, principalmente fenótipos com suspeita de segregação mendeliana	2019	144	2	0	2165
Total	16151	963	56	63	1723

As estratégias para a resolução desses problemas começam desde o planejamento do estudo, como o uso de projetos de dois estágios, em que uma parte da amostra é testada com todos os marcadores e a outra é testada apenas com os marcadores com resultados significativos no primeiro estágio. Outras estratégias envolvem confirmação em outras amostras e metanálises. É sempre bom lembrar que resultados obtidos em um experimento de ligação devem ser confirmados com outro tipo de estudo ou em mais de uma amostra.

O método *lod score*

O método *lod score* foi proposto por Morton em 1955 e foi uma contribuição muito importante e duradoura desse pesquisador para a genética humana. *Lod score* vem

do inglês *logarithm of an odds ratio*. Sendo *odd ratio* definida como a razão entre a probabilidade de um genótipo específico ser observado em determinada família, em uma determinada fração de recombinação diferente de 0,5 contra a mesma probabilidade, computada considerando segregação independente. O logaritmo é aplicado por ser mais conveniente, já que traz os valores a uma escala mais adequada, assim como permite que os valores obtidos em diversas famílias possam ser somados entre si. Quanto maior o *lod score*, maior a evidência a favor de ligação, e, sem levar em conta múltiplos testes e outras complicações, *lod scores* com valores acima de três são altamente indicativos de ligação. O fato de os *lod scores* poderem ser somados é especialmente interessante quando se pretende realizar uma metanálise. Com o algoritmo que permite o cálculo do

lod score em *pedigrees* mais complexos proposto por Elston e Stewart (1971) e posteriormente com o programa de Ott (1974), que permitia a realização desses cálculos de maneira mais amistosa, o método *lod score* ganhou popularidade e vem sendo amplamente utilizado.

Outros métodos de estudo de ligação gênica

O livro de Jurg Ott, *Analysis of Human Genetic Linkage*, apresenta uma revisão extensa e abrangente sobre ligação e os métodos empregados para o seu estudo. O método *lod score* vem sofrendo diversas modificações e aperfeiçoamentos que tentam levar em conta as particularidades das populações e características estudadas.

De forma geral, os métodos de análise de ligação consideram a proporção de alelos em marcadores compartilhados entre parentes devido à descendência (em inglês, *IBD* – *identity by descend*) contra a proporção dos alelos compartilhados devido ao acaso (em inglês, *IBS* – *identity by state*) e a concordância da característica estudada com essas proporções.

Partindo de mapas de *IBD*, os métodos de análise de ligação podem ser divididos em dois grandes grupos: análise paramétrica e análise não-paramétrica. A diferença principal entre os dois tipos de análise é que na análise paramétrica, baseada principalmente no *lod score*, é necessário estabelecer um modelo de herança para a característica em estudo, ao passo que na análise não-paramétrica esse modelo não é necessário. As vantagens e desvantagens de cada método são motivos de discussão entre especialistas da área e dependem muito da amostra a ser estudada, bem como da característica. De maneira breve, pode-se dizer que os métodos paramétricos têm maior poder estatístico quando utilizado em características com herança bem determinada. Em características complexas, frequentemente os métodos não-paramétricos são

preferidos. De qualquer maneira a escolha do método depende de muitos outros fatores além dos descritos aqui.

ASSOCIAÇÃO

Estudos de associação visam a identificar um marcador que tenha relação com um fenótipo no nível populacional. Isso se deve a um possível desequilíbrio de ligação entre dois genes. Dizemos que dois genes estão em desequilíbrio de ligação quando as frequências simultâneas dos alelos forem diferentes daquelas esperadas por acaso. Quanto mais próximos estiverem os alelos no cromossomo, menor será a probabilidade de recombinação e maior será a probabilidade dos dois genes estarem em desequilíbrio de ligação.

Isso se deve ao fato de que quanto mais próxima for a ligação, maior será o tempo que o desequilíbrio permanecerá na população. Em uma amostra em que o casamento é ao acaso, a tendência ao equilíbrio é igual a uma parte da taxa de recombinação (θ); e decresce um valor igual a θ em cada geração. Em *loci* não-ligados, a frequência dos haplótipos percorre metade do caminho para o equilíbrio a cada geração; quando $\theta = 0,10$, serão necessárias sete gerações; quando $\theta = 0,01$, serão 69 gerações; e quando $\theta = 0,001$, serão necessárias 693 gerações. Isto é, quanto menor a taxa de recombinação (θ), maior a probabilidade de dois genes estarem em desequilíbrio de ligação e um número maior de gerações será necessário para população chegar ao equilíbrio. Com isso, as perspectivas para se detectar um desequilíbrio de ligação são menores para mutações mais antigas do que para mutações mais recentes, porém outros fatores evolucionários como deriva genética, migração e mistura, por exemplo, podem amplificar a visibilidade do desequilíbrio ou mesmo mimetizá-lo (Rao e Province, 2001).

Muitos estudos de associação com doenças infecciosas são feitos com a utilização de

SNP (*single nucleotide polymorphism*), STR (*short tandem repeats*) e marcadores clássicos de sistemas sanguíneos (como ABO, Rh e Duffy, por exemplo). Nesses estudos são usados métodos do tipo caso–controle, nos quais indivíduos doentes são comparados com indivíduos saudáveis pertencentes à mesma população, e testes do tipo TDT (*transmission/disequilibrium-test*), nos quais é analisada a transmissão dos alelos entre os indivíduos afetados e não afetados.

A forma mais utilizada para estudar associação, devido à facilidade na obtenção de amostras apropriadas, é por meio da utilização de testes do tipo caso–controle, que se dá pela genotipagem dos indivíduos dentro de uma população. Compara-se o marcador ou as frequências genotípicas de STR ou SNP nos indivíduos que apresentam a característica em estudo (afetados) com as frequências dos indivíduos do restante da amostra (não afetados). Se o desequilíbrio de ligação entre os marcadores for significativamente diferente entre caso e controle, está caracterizada uma associação entre marcador e fenótipo (Nilesen e Weir, 2001).

Em 2003, estudos desse tipo revelaram que existe uma associação entre indivíduos do grupo sanguíneo Duffy negativo e o número de episódios de malárias que o grupo sofreu em Portuchuelo, RO (Beiguelman et al, 2003). Nesse trabalho, uma população composta por 182 indivíduos foi analisada. Testes com marcadores de grupos sanguíneos realizados entre indivíduos infectados e indivíduos não infectados comprovaram a associação entre o grupo sanguíneo Duffy negativo e a resistência pela infecção por *P. vivax*, fato já conhecido e presente na literatura científica.

Para o teste TDT é necessária a informação genotípica e trios de indivíduos – pai, mãe e filho – e pelo menos um dos pais deve ser heterozigoto para que o teste seja informativo. A grande vantagem desse tipo de teste, que vem ganhando cada vez mais importância, é

que ele não é influenciado pela estratificação da amostra e nem por casos de mistura étnica. Pesquisadores têm contribuído bastante para estender o teste TDT para outras situações, como marcadores com mais de um alelo (Sham e Curtis, 1995; Bickeboeller e Clerget-Darpoux, 1995; Rice et al., 1995), casos em que se tem apenas um dos pais para o estudo (Sun et al., 1999), casos em que não se tem nenhum dado dos pais usando pares de irmãos (Spielman e Ewens, 1998; Hovarth e Laird, 1998; Boehnke e Langefeld, 1998), o que traz informações sobre alelos que não segregam em indivíduos afetados.

Tanto no teste de caso–controle quanto no teste TDT, o que deve ser analisado é o marcador depois de a população estar corretamente organizada por fenótipo ou positividade quanto à infecção da doença em questão. A principal barreira nos estudos de associação é que não existe garantia da presença de um desequilíbrio de ligação na região de interesse; sendo assim, se não há desequilíbrio de ligação, os testes de associação não terão alto poder estatístico, fato que torna ainda mais importante a escolha da população a ser estudada (Rao e Province, 2001).

Um dos problemas nos estudos de associação surge quando se deve fazer a análise com marcadores multialélicos. Alguns métodos, como aumentar o número de classes, acarretam aumento do número de graus de liberdade e, assim, perda de poder estatístico. Outro mecanismo seria diminuir o número de alelos, o que pode levar a um problema estatístico quando se tenta comparar todos os alelos possíveis. O melhor modelo seria baseado na verossimilhança para os testes de desequilíbrio de ligação que pode utilizar vários alelos (Terwilliger, 1995); o teste torna-se mais conservativo à medida que se aumenta o número de alelos sem perder o poder estatístico (Rao e Province, 2001).

Quando se examina um gene candidato, vários sítios polimórficos podem existir entre

ou próximos ao gene, do qual uma parte pode ser escolhida para a genotipagem. Cada um desses sítios pode ser funcional ou pode estar associado a outro sítio funcional. Examinando individualmente SNP dessa região não se obtém um grande poder para detecção de associações quando se compara com o uso de haplótipos (Rao e Province, 2001).

Uma importante fonte de estudos de associação são os haplótipos, isto é, segmentos dos cromossomos que são transmitidos através das gerações como regiões de seqüências de DNA que são compartilhados por muitos indivíduos sem sofrer recombinação. Sendo assim, permitem a busca de genes candidatos associados a doenças e outras características relevantes.

Para isso alguns geneticistas utilizam o HapMap, que é um catálogo de variantes genéticas comuns que ocorrem nos seres humanos. Nesse catálogo podem ser encontradas as descrições das variantes, o local onde ocorrem e como estão distribuídas nos indivíduos dentro das populações e entre as populações em diferentes partes do mundo. O Projeto HapMap tem o objetivo de fornecer informações a respeito da relação de genes com o possível risco a algumas doenças para que outros pesquisadores possam concluir novas descobertas e, assim, proporcionar novos métodos de tratamento, diagnóstico e prevenção de doenças (*International HapMap Project*).

Procurou-se nos tópicos anteriormente resumidos dar uma idéia das ferramentas que os modernos programas computacionais empregam no tratamento de características complexas e na separação de mecanismos monogênicos que podem afetar essas características.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Bateson W. G. Mendel: Experiments in Plant Hybridisation. J Royal Hort Soc 1901; 26:1-32.
2. Bateson W. Hybridisation and cross-breeding as a method of scientific investigation. Journal of the Royal Horticultural Society 1900; 24:59-66.
3. Bateson W, Punnett RC. On the inter-relations of genetic factors, Vol 6, Number 1 / December, 1911 http://www.springerlink.com/index/L80H71270_0246G85.pdf Bateson, W, Punnett, RC. On the Inter-relations of Genetic Factors. Proceedings of the Royal Society, 84, B. 1911, S. 3
4. Beiguelman B, Alves FP, Moura MM, Engracia V, Nunes ACS, Heckmann MIO et al. The association of genetic markers and malaria infection in the Brazilian Western Amazonian region. Mem Inst Oswaldo Cruz 2003; 98. [cited 2006 Nov 17]. Available at: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762003000400004&lng=en&nrm=iso>.
5. Bickeboeller H, Clerget-Darpoux F. Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers. Genet Epidemiol 1995; 12:865-70.
6. Boehnke M, Langeveld CD. Genetic association base don discordant sib pairs: The discordant-allelic test. Am J Hum Genet 1998; 62:950-61.
7. Hovarth S, Laird MN. A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet 1998; 63:1886-97.
8. International HapMap Project. [cited 2006 nov 10]. Available at: <http://www.hapmap.org/index.html>.
9. Li L, Wang QK, Rao S, LINKAGE programs: linkage analysis for monogenic cardiovascular diseases. Methods Mol Med. 2006; 128:41-60.
10. Mendel G. Versuch über Pflazen-Hybriden. Verh Naturforsch Ver Brünn 1866; 4:3-47.
11. Nielsen DM, Weir BS. Association studies under general disease models. Theoretical population biology 2001; 60:253-63.
12. Rao DC, Province MA. Genetic dissection of complex traits. In: Borecki IB, Suarez BK. *Linkage and associaton: basic concepts*. San Diego: Academic Press, 2001. p. 45-66.
13. Rice JP, Neuman RJ, Hoshaw SL, Daw EW, Gu C. TDT with covariates and genomic screen with mod scores: their behavior on simulated data. Genet Epidemiol 1995; 12:659-64.

14. Sham PC, Curtis D. An extended transmission/disequilibrium test (TDT) for multiple allele marker loci. *Ann Hum Genet* 1995; 59: 323-36.
15. Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am J Hum Genet* 1998; 62:450-8.
16. Stewart WC, Thompson EA. Improving estimates of genetic maps: a maximum likelihood approach. *Biometrics* 2006 Sep;62(3):728-34.
17. Sun F, Flanders WD, Yang Q, Khoury MJ. Transmission disequilibrium test (TDT) when only one parent is available: the 1 – TDT. *Am J Epidemiol* 1999; 150:97-104.
18. Tapper W, Collins A, Gibson J, Maniatis N, Ennis S, Morton NE. A map of the human genome in linkage disequilibrium units. *Proc Natl Acad Sci USA* 2005 Aug 16; 102(33): 11835-9. Epub 2005 Aug 9.
19. Terwilliger JD. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J Hum Genet.* 1995; 65:777-87.