

**Estudos de simetria  
na associação genética  
usando dados de trios**

Maria Jacqueline Batista

TESE APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO TÍTULO  
DE  
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientadora: Profa. Dra. Júlia Maria Pavan Soler

Durante o desenvolvimento deste trabalho a autora recebeu apoio financeiro da  
CAPES e FAPESP processo N° 06/53612-0

São Paulo, dezembro de 2011

**Estudos de simetria na associação  
genética usando dados de trios**

Esta tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa realizada por Maria Jacqueline Batista em 02/12/2011.

O original encontra-se disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Profa. Júlia Maria Pavan Soler (Orientadora) - IME/USP
- Prof. Carlos Alberto de Bragança Pereira - IME/USP
- Prof. Dalton Francisco de Andrade - INE/CTC/UFSC
- Profa. Clarice Garcia Borges Demétrio - ESALQ/USP
- Profa. Suely Ruiz Giolo - DEST/UFPR

*“Aleluia!*

*Louvai, ó servos do Senhor, louvai o nome do Senhor.*

*Bendito seja o nome do Senhor, agora e para sempre.*

*Desde o nascer ao pôr-do-sol, seja louvado o nome do Senhor.*

*O Senhor é excelso sobre todos os povos,  
sua glória ultrapassa a altura dos céus.”*

Salmos 112, 1-4

*“Jesus olhou para eles e disse:  
aos homens isto é impossível,  
mas a Deus tudo é possível.”*

Mateus 19,26

*“O coração do homem  
dispõe o seu caminho,  
mas é o Senhor que  
dirige seus passos.”*

Provérbios 16,9

*“Louvai o Senhor,  
porque Ele é bom,  
cantai à glória de Seu nome,  
porque Ele é amável.”*

Salmos 134,3

*“Este é o dia que o Senhor fez: seja para nós dia de alegria e de felicidade.*

*Senhor, dai-nos a salvação; dai-nos a prosperidade, ó Senhor!*

*Bendito seja o que vem em nome do Senhor!*

*Da casa do Senhor nós vos bendizemos.*

*O Senhor é nosso Deus, ele fez brilhar sobre nós a sua luz.*

*Organizai uma festa com profusão de coroas. E cheguem até os ângulos do altar.*

*Sois o meu Deus, venho agradecer-vos. Venho glorificar-vos, sois o meu Deus.*

*Dai graças ao Senhor porque ele é bom, eterna é sua misericórdia.”*

Salmos 117, 24-29

*A Deus,  
minha filha Ana Yasmin, meu querido marido Juvêncio,  
meus pais: Otacílio e Socorro, meu irmão Júlio,  
e minha avozinha querida Ana M. Batista (In memoriam).*

# Agradecimentos

Agradeço,

A Deus Todo Poderoso, pela saúde e oportunidade. À Mãe Imaculada pelas graças alcançadas.

À minha família, alicerce de tudo: minha mãe Socorro, meu pai Otacílio e meu irmão Júlio. À minha avozinha: Ana, por toda dedicação e amor (muitas saudades). Amo vocês.

Ao meu amado marido Juvêncio, por tudo: paciência, amor, carinho, conselhos... E por neste doutorado ter nascido nossa filha Ana Yasmin, amo muito vocês dois! Meu nego, saiba que a nossa família é a maior alegria da minha vida. Porque família é tudo.

A D. Gracilene, pela força, e por ficar com minha filha enquanto eu tinha que viajar para resolver as pendências desta tese.

À minha orientadora, profa. Júlia Maria, sou muito agradecida a ela, não somente por ter me guiado neste tema e dado energia na orientação deste trabalho, mas também por ser uma amiga em todos os momentos desta trajetória, foi muito bom conhecê-la e aprender com ela.

Aos professores do IME-USP, em especial, Julio Singer, Elisabeti Kira e Antônio Carlos e os do DEMA-UFC, em especial Ana Maria, Maurício Mota, André Shiguemoto, Júlio Barros, Rosa Mota, Sílvia Freitas, João Welliandre e Ronald Nojosa (agradeço em especial as palavras de incentivo e implementação computacional que foi de GRANDE ajuda, agradeço também a sua esposa Francilene pelo apoio), e também as meninas da secretaria, Margeri e Luisa.

Aos meus amigos do IME-USP, em especial, a Michelli e Horácio, Tatiana e Alessandro, Patrícia e Raydonal, Lane e Marcelo, Luz Marina, Rafael, Michel, Caio, Alexandre, Gleiciane, Tatiana, Gérman e Lizandra.

À Núbia que me ajudou em todas as fases deste trabalho, com palavras, hospedagem, programas computacionais, ou seja, ela faz parte deste trabalho, muito obrigada!

Ao Laboratório de Genética e Cardiologia Molecular do Instituto do Coração da Faculdade de Medicina da Universidade de São Paulo (InCor-USP), pelos dados reais, em especial aos Drs. José Eduardo Krieger e Alexandre C. Pereira.

À banca examinadora, prof. Carlos Alberto de Bragança Pereira, prof. Dalton Francisco de Andrade, profa. Clarice Garcia Borges Demétrio e em especial a profa. Suely Giolo, pelo apoio e INCENTIVO no decorrer deste trabalho.

À FAPESP e CAPES pelo auxílio financeiro.

Não dá para citar todos os nomes que merecem agradecimentos, pois são muitas pessoas, mas saibam que todas estão no meu coração ♡.



# Resumo

## Estudos de simetria na associação genética usando dados de trios

O grande desafio da Epidemiologia Genética, atualmente, é identificar, em um espaço de variáveis preditoras de alta dimensão e esparsas, fatores de risco genéticos para doenças complexas. Um delineamento amostral útil nestes estudos é coletar dados de trios, que são pequenos núcleos familiares (pai e mãe, livres da doença, e filho afetado) e, em cada indivíduo, obter dados do genótipo de marcadores moleculares, sendo a plataforma de marcadores do tipo SNPs (do inglês, *Single Nucleotide Polymorphism*), com cerca de 1 milhão de variáveis preditoras genéticas, a mais adotada. Neste trabalho é proposto um procedimento em múltiplos estágios para identificar SNPs associados com a doença em dados de trios. A primeira etapa do procedimento é baseada em uma série de análises unilocus (para cada variável preditora), usando um teste de simetria em tabelas de contingência  $2 \times 2$  (conhecido, em Genética, como teste TDT, do inglês, *Transmission Disequilibrium Test*). Em um segundo estágio da análise, os resultados destes testes são usados para construir uma estatística de somas acumuladas padronizadas (CUSUM) que permite a seleção de conjuntos de SNPs (isto é, conjuntos de variáveis preditoras), possivelmente associados com a doença. Como um terceiro passo da análise, nas regiões selecionadas no passo dois, são realizadas análises de simetria via testes exatos considerando tabelas  $2 \times 2$  e  $4 \times 4$  (pares de SNPs). A formulação do TDT em termos de testes de simetria é uma inovação na área de Genética e facilita a extensão do caso unilocus para o multilocus. A contribuição deste trabalho reside ainda na formulação exata do teste que é útil em situações de amostras pequenas que ocorrem com frequência em dados de trios. Neste caso inferências parciais foram realizadas a partir de decomposições apropriadas da função de verossimilhança. A modelagem do problema em termos do modelo logístico permitiu concluir que não é necessário corrigir a associação para o efeito de covariáveis avaliadas nos pais. O procedimento é implementado usando recursos dos aplicativos PLINK e R. A aplicação é realizada utilizando dados de 71 trios da população brasileira, em que os indivíduos caso (filhos) foram definidos em termos da ocorrência de uma cardiopatia e, em cada um dos 213 indivíduos, estão disponíveis dados genéticos de uma plataforma de SNPs.

**Palavras-chave:** Mapeamento genético, Análise multilocus, Estudos de simetria, Dados de trios, Teste TDT.

# Abstract

## Symmetry studies in the genetic association using data from trios

Currently, the great challenge of Genetic Epidemiology is to identify, in a high dimensional and sparse space of predictor variables, genetic risk factors for complex diseases. A useful sampling design in these studies is to collect data from trios, which are small nuclear families (father and mother, free from disease, and affected child), and obtain genotypic information from each individual. The molecular markers platform most commonly used for this purpose is of SNPs (Single Nucleotide Polymorphisms), with about 1 million genetic predictor variables. This work proposes a multi-stage procedure to identify SNPs associated with disease using data from trios. The first step of the procedure is based on a series of single locus analysis (for each predictor variable) using a test for symmetry in  $2 \times 2$  contingency tables (known in genetics as TDT (Transmission Disequilibrium Test)). In a second stage of the analysis, the results of these tests are used to construct a standard statistic of the cumulative sums (CUSUM), which allows the selection of sets of adjacent SNPs (ie, sets of predictor variables), possibly associated with the disease. As a third step of the analysis, in the regions selected in step two, are performed an extended analysis of symmetry considering  $4 \times 4$  contingency tables. The TDT formulation in terms of symmetry tests is an innovation in the genetics area and facilitates the extension of the single locus analysis to the multiloci case. The contribution of this work lies in the exact formulation of the symmetry test for square contingency tables that is useful in situations of small sample sizes that often occur in data from trios. In this case, partial inferences were performed from appropriate decompositions of the likelihood function. The structural modeling of the problem in terms of logistic model allowed us to conclude that there is no need to adjust the association for data from parents, but only for the effect of covariates evaluated in each parental haplotype. The procedure is implemented using resources of the R statistical environment and Plink. The application is performed using real data from 71 trios of the Southeast Brazilian population, in which affected child was defined in terms of the occurrence of one congenital heart disease, and in each of the 213 individuals, genomic data were collected using Affymetrix SNP 6.0 platform.

**Keywords:** Genetic mapping, Multiloci analysis, Symmetry studies, Data trios, Test TDT.

## Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

