



## Defesa de Exame de Qualificação ao Doutorado em Modelagem Computacional

**DATA:** 31/03/2014  
**HORÁRIO:** 14h  
**LOCAL:** Anfiteatro 03 – Prédio Engenheiro Itamar Franco  
(Faculdade de Engenharia)

### **“Desenvolvimento de métodos, técnicas e ferramentas para estudos de associação em escala genômica”**

**Doutorando:** Fabrizzio Condé de Oliveira  
**Orientador:** Prof. Carlos Cristiano Hasenclever Borges

#### **Banca Examinadora:**

Prof. Henrique Steinhert Hippert – UFJF (Presidente), D.Sc.  
Prof. Priscila Vanessa Zabala Capriles Goliatt, UFJF, D.Sc  
Prof. Marcos Vinicius Gualberto Barbosa da Silva – EMBRAPA, D.Sc.

#### **Resumo:**

Estudos de associação em escala genômica (*Genome-Wide Association Studies* – GWAS) têm demonstrado ser uma estratégia poderosa e com grande potencial na descoberta de genes relacionados com determinada característica fenotípica. Métodos comumente em GWAS avaliam separadamente o impacto de cada marcador do tipo SNP (*single nucleotide polymorphism*) no fenótipo considerado. Essa avaliação pode ser feita por meio de regressão linear simples para o caso de fenótipo contínuo, ou por meio de regressão logística para fenótipo dicotômico mapeado por caso-controle. Na sua aplicação mais usual, o GWAS inicia calculando o p-valor do teste estatístico usado na associação entre o SNP e a característica, através do qual ordena-se de forma crescente os marcadores. Quanto menor o p-valor, mais significativo estatisticamente é o marcador com relação ao fenótipo. Em diversos trabalhos de GWAS em humanos, esse procedimento encontrou SNPs que marcam genes que atuam diretamente no risco de desenvolvimento de determinada doença como Alzheimer, doença de Crohn, diabetes dos tipos 1 e 2 entre outras.

Um possível cenário é quando o fenótipo é determinado pela influência de vários genes com pequenos efeitos aditivos sobre o mesmo. Assim, pode-se não identificar os marcadores relevantes, pois separadamente seus efeitos possuem baixa significância. Outra situação é quando dois ou mais SNPs, pouco significativos isoladamente, demonstram grande efeito de interação, gerando padrões complexos na relação genótipo-fenótipo (epistasia). Podem ocorrer os dois casos anteriores simultaneamente, dificultando ainda mais a detecção correta de SNPs. Outro ponto negativo desse modelo monoatributo, é demonstrar baixo poder de explicação para a variância do fenótipo em problemas de regressão. Atualmente, os *chips* de alta densidade possuem aproximadamente 800.000 marcadores genéticos e o número de indivíduos nas amostras consideradas em GWAS estão na ordem de centenas ou poucos milhares, gerando uma limitação para aplicação de diversas técnicas multivariadas. Desta forma, para possibilitar a seleção simultânea de vários SNPs, onde há o desequilíbrio entre o tamanho da amostra e a quantidade de marcadores no conjunto de dados, pode-se usar as técnicas de Máquina de Vetores Suporte (SVM - *Support Vector Machine*) para problemas de classificação e para problemas de regressão (SVR - *Support Vector Regression*). O SVM e o SVR têm demonstrado potencial para GWAS em humanos, outros animais e em vegetais. O presente trabalho tem por objetivo propor um novo método baseado em um modelo multiatributo, construído a partir de filtros estatísticos tradicionais em GWAS juntamente com SVM ou SVR, objetivando a captura simultânea de efeitos aditivos e/ou não-lineares dos marcadores sobre o fenótipo. Para eliminar possíveis marcadores redundantes, o SVM/SVR é encapsulado dentro de um algoritmo genético para a seleção dos marcadores do tipo SNP que expliquem melhor a característica fenotípica.