



UNIVERSIDADE FEDERAL DE PERNAMBUCO - UFPE  
CENTRO DE INFORMÁTICA - CIn  
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**VISUALIZAÇÃO DE REDES REGULATÓRIAS A  
PARTIR DE MODELOS DE REGRESSÃO E  
SELEÇÃO DE ATRIBUTOS**

**Proposta de Trabalho de Graduação**

**Aluno:** Ruan Vasconcelos Bezerra Carvalho (rvbc)

**Proponente:** Ivan Gesteira Costa Filho (igcf)

Recife, 14 de abril de 2011

# Sumário

1	Contexto	3
2	Objetivo	3
3	Cronograma	3
4	Possível Avaliador	4
5	Assinaturas	6

# 1 Contexto

Nos últimos anos, um grande número de avanços técnicos experimentais tem nos permitido obter uma grande quantidade de informação sobre sistemas vivos em uma escala genômica. Embora os genomas completos de muitos organismos estejam disponíveis graças ao progresso feito na tecnologia de sequenciamento, o desafio é entender como os genes são regulados individualmente dentro do interior da célula [Babu 2008].

A identificação de sinais de regulação pode ser feita, por exemplo, com algoritmos combinatórios determinísticos baseados na contagem de frequência de palavras nas regiões promotoras dos genes. Esses métodos se baseiam na busca exaustiva de regiões promotoras por seqüências de nucleotídeos de diversos tamanhos e avaliam a significância das seqüências obtidas através de um método estatístico [Kelles 2002].

Uma estratégia normalmente usada em biologia computacional é a busca de grupos de genes que exibem padrões similares de expressão através de condições experimentais. Dado que genes com mesmo padrão de expressão devem ser controlados pelos mesmos mecanismos regulatórios, uma maneira de inferência de redes regulatórias é a análise conjunta de dados de expressão gênica e de regiões promotoras [Kelles 2002]. Por exemplo, temos o uso de regressão linear, em que afinidades de modificações de histonas são as variáveis explicativas e os valores de expressão gênica em diversas condições biológicas são as variáveis explicadas. Neste contexto, os coeficientes da regressão estimados são indicadores de quais modificações de histonas são responsáveis pela expressão de quais genes em qual condição biológica [Kelles 2002].

# 2 Objetivo

O objetivo deste trabalho é a utilização de modelos de mistura para a estimação de vários modelos de regressão, juntamente com a criação de funcionalidades que permitam a visualização dos mesmos, a fim de facilitar a interpretação dos resultados biológicos. Para isso, utilizaremos dados de expressão gênica do desenvolvimento de células do sistema linfático [Costa 2010], os quais já foram analisados pelo proponente e são de domínio público. O software será implementado na linguagem de programação Python e usará uma plataforma de código aberto [Pymix].

# 3 Cronograma

Atividade	Março		Abril				Maio				Junho			
Levantamento bibliográfico	X	X	X	X	X	X								
Implementação do método proposto			X	X	X	X								
Execução dos experimentos					X	X	X	X						
Escrita do relatório					X	X	X	X	X					
Elaboração da apresentação											X	X	X	X

## 4 Possível Avaliador

Ricardo Bastos Cavalcante Prudêncio

## Referências

- [Babu 2008] Babu, M.M. (2008) Computational approaches to study transcriptional regulation. *Biochemical Society Transactions*, 36: 758-765
- [Kelles 2002] Keles, S. et al. (2002) Identification of regulatory elements using a feature selection method. *Bioinformatics*, 18(9): 1167-1175.
- [Costa 2010] I. G. Costa, H. Roeder, F. A. T. de Carvalho, Predicting Gene Expression in T cell Differentiation from Histone Modifications and Transcription Factor Binding Affinities by Linear Mixture Models, *Bioinformatics*, submetido, 2010.
- [Pymix] <http://algorithmics.molgen.mpg.de/Software/PyMix>

## 5 Assinaturas

---

Ivan Gesteira Costa Filho

---

Ruan Vasconcelos Bezerra Carvalho