



ciência desenvolvimento sociedade
**XXVI SALÃO DE
INICIAÇÃO CIENTÍFICA**

20 a 24 de outubro - Campus do Vale - UFRGS



Evento	Salão UFRGS 2014: SIC - XXVI SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2014
Local	Porto Alegre
Título	O Transcriptogramer
Autor	JOÃO MEDEIROS DINIS
Orientador	RITA MARIA CUNHA DE ALMEIDA

A análise quantitativa da expressão gênica de um organismo caracteriza o estado em que o mesmo se encontra, e uma técnica muito utilizada para esta medida é o Microarranjo, mas esta é considerada demasiadamente ruidosa. Com base em toda a informação acerca do funcionamento celular disponível na rede, desenvolveu-se um método capaz de reduzir o ruído dessa medida. O Transcriptograma é um método para apresentar, analisar e tratar dados de transcrição de genoma completo que reduz o ruído e facilita a interpretação biológica destes. Inicialmente uma lista ordenada de genes é produzida, a partir da rede de associação proteína-proteína disponível no sítio STRING, de tal modo que a probabilidade de que os genes que são funcionalmente associados decai exponencialmente com o distanciamento na lista. Este ordenamento apresenta justificativa biológica, evidenciada pela segregação de diferentes funções biológicas do Gene Ontology e rotas metabólicas do KEGG. Perfis de Transcriptograma são obtidos pela média da expressão gênica tomada sobre genes vizinhos nesta lista, de forma similar ao método de amplificação de sinal *Boxcar Average*. Além de aumentar a precisão das medições de expressão dos conjuntos de genes funcionalmente relacionados, o Transcriptograma também melhora a reprodutibilidade das medidas. Na primeira parte do projeto foi feito um estudo sobre o método para a sua melhor compreensão; investigou-se métodos computacionais para obtenção de melhores resultados e os conceitos e fundamentos básicos de biofísica e estatística para melhor compreensão destes. Na etapa seguinte analisou-se a qualidade do Transcriptograma aplicado ao tratamento de dados para uso em diagnósticos. Foram realizados diagnósticos de Psoríase, Esclerose Múltipla e estágios desta. Em cada diagnóstico, utilizou-se o método de *Machine Learning* para definir os melhores identificadores de cada classe a partir de amostras de treinamento; a partir destes, outro conjunto de amostras foi classificado com os métodos LDA, QDA, dLDA, Naive Bayes, entre outros conhecidos da literatura. Realizou-se uma varredura de parâmetros para verificar sob quais condições o Transcriptograma mostra-se efetivo na melhora do diagnóstico destas doenças. Para a realização deste estudo, foram utilizadas as linguagens de programação C, Python e R. Os resultados obtidos foram muito positivos, aumentando consideravelmente o índice de acerto dos diagnósticos, e artigos acadêmicos baseados nestes trabalhos estão sendo redigidos e submetidos a revistas de alta importância científica na área.