



Evento	Salão UFRGS 2019: FEIRA DE INOVAÇÃO TECNOLÓGICA DA UFRGS - FINOVA
Ano	2019
Local	Campus do Vale - UFRGS
Título	Análise de Transcriptograma Para Dados de Sarcoma de Ewing
Autores	MAURICIO GOMES DE QUEIROZ MATHEUS GIBEKE SIQUEIRA DALMOLIN MARIALVA SINIGAGLIA
Orientador	RITA MARIA CUNHA DE ALMEIDA

RESUMO DE ATIVIDADES

TÍTULO DO PROJETO: Análise de Transcriptograma Para Dados de Sarcoma de Ewing

Aluno: Maurício Gomes de Queiroz

Orientadora: Rita Maria Cunha de Almeida

Coorientação: Marialva Sinigaglia, Instituto do Câncer Infantil

A técnica do Transcriptograma ^[1] baseia-se na produção de perfis de expressão gênica a partir de medidas de expressão de genoma inteiro, tais como as obtidas por micro arranjos ou RNA-Seq.

Os perfis consistem na projeção dos dados de expressão sobre uma lista de genes ordenadas de tal maneira que os genes pertencentes a uma mesma função biológicas estão localizados em regiões próximas na lista. A lista ordenada é feita com base na informação sobre interação proteína-proteína, como disponibilizado na base de dados STRING ^[2], por meio de uma simulação de Monte Carlo que minimiza as distâncias na lista de produtos gênicos que participam de uma mesma rota metabólica. Neste trabalho, utilizamos a lista ordenada para *Homo sapiens* como obtida por de Almeida e colaboradores ^[3]. A lista ordena 9684 genes, cujos produtos gênicos foram listados como associados a pelo menos um outro produto gênico com escore de confiança 800.

Com a lista ordenada dados relativos à expressão gênica de cada amostra são projetados de tal maneira que a expressão de cada gene é associada à posição do mesmo e então uma média sobre as expressões dos genes de cada posição e em posições vizinhas de raio R à esquerda e à direita é atribuída ao gene central. A influência de diversos valores para R foi estudada. Esta média janelada reduz o ruído inerente à técnica de medidas de expressão gênica enquanto que a lista ordenada pelas funções biológicas possibilita que a razão sinal-ruído seja melhorada, aumentando o poder do teste.

Neste projeto, a técnica Transcriptogramer foi estudada e aplicada no contexto do sarcoma de Ewing. Queremos determinar diferenças na expressão dos pacientes que obtiveram um melhor prognóstico, e portanto, após considerar diferentes conjuntos de dados para serem analisados, escolhemos três experimentos cujos metadados foram disponibilizados juntos com os dados de expressão gênica. Os dados utilizados foram retirados do GEO - Gene Expression Omnibus repository ^[4].

Os dados passaram por um processo de controle de qualidade, e foram normalizados pela técnica de Robust Micro Average (RMA). Após normalizados, foi feito um transcriptograma dos dados. Regiões diferencialmente expressas foram estudadas mais a fundo com um enriquecimento funcional, do qual um conjunto de genes diferencialmente expressos de interesse foi obtido. Diversas comparações foram feitas utilizando o Transcriptogramer, e seus resultados discutidos.

Utilizando agora outras técnicas, como o método Kaplan-Meier, os genes que se mostraram significativos são validados. As mesmas validações são feitas para os três conjuntos de dados, que ao final, indicam um conjunto de genes cuja expressão pode indicar o prognóstico da doença.

[1]: (RYBARCZYK-FILHO, CASTRO, *et al.*, 2011) (DA SILVA, PERRONE, *et al.*, 2014)

[2]: (SZKLARCZYK, MORRIS, *et al.*, 2017)

[3]: (DE ALMEIDA, CLENDENON, *et al.*, 2016)

[4]: (<https://www.ncbi.nlm.nih.gov/geo/>)