VOZES DIVERSAS
DIFERENTES SABERES

CONHECIMENTO FORMAÇÃO INOVAÇÃO
Salão UFRGS 2018

SALÃO DE INICIACAO CIENTIFICA
XXX SIC

15 A 19 OUTUBRO
CAMPUS DO VALE

UFRGS
PROPESQ

Ciclo Celular Detalhado pela Análise de Componentes Principais

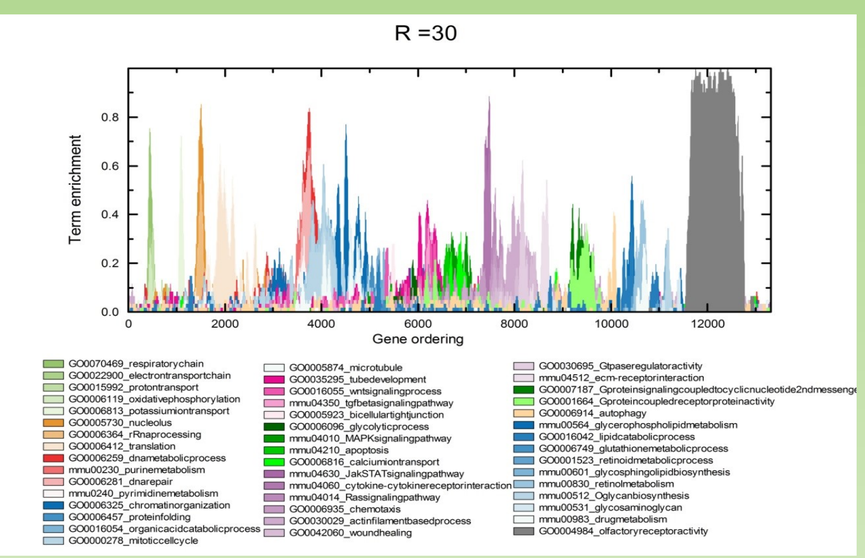
Aluno: Lars Leonardo Sanhudo de Souza
Orientadora: Prof. Dr. Rita Maria Cunha de Almeida
Departamento de Física / UFRGS

INTRODUÇÃO

As células apresentam diferentes estados metabólicos, que podem ser caracterizados pelo perfil de sua expressão gênica, cuja medida pode ser feita por meio de técnicas de micro-arranjos ou RNASeq (sequenciamento de RNA) que resultam num perfil das quantidades relativas de RNA mensageiro produzido em um dado instante de tempo pela célula. Este medida estima o perfil do conjunto de proteínas atuantes na célula.

O ciclo celular é uma sucessão de estados metabólicos da célula, sendo composto de várias fases: o crescimento da célula (conhecido como fase G1), a duplicação do seu DNA, conhecida como síntese (fase S) e por fim, a preparação (G2) e realização da mitose (M).

O método do transcriptograma [1] consiste em calcular médias sobre o valor da expressão dos genes vizinhos em uma lista ordenada segundo a função biológica dos genes.



O numero de vizinhos escolhidos para a realização da média é chamado de raio do transcriptograma. A utilização das médias da expressão gênica é interessante por minimizar a intensidade do ruído, preservando o sinal daquele grupo de genes, melhorando portanto a razão sinal-ruído.

A análise das componentes principais (PCA) é um método que analisa os dados visando a redução de variáveis, identificando as relevantes ao problema. Neste caso, cada expressão gênica é considerada um eixo de um espaço de muitas dimensões. A PCA identifica as direções neste espaço que contêm a maior variação dos dados. A utilização de transcriptogramas como dados de entrada para PCA resulta em perfis de transcrição do genoma inteiro como as componentes do PCA. Assim, não somente a variação de genes individuais, mas a correlação entre expressões gênicas são identificadas variando ao longo do ciclo celular.

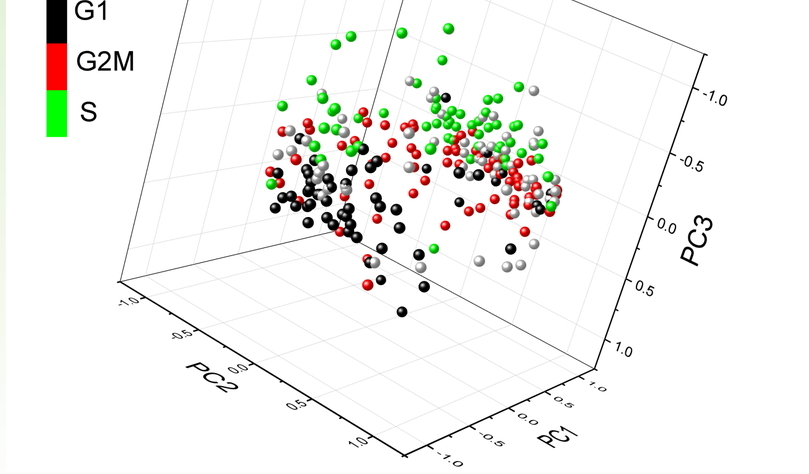
A análise conjunta de transcriptograma e PCA, como mostram nossos resultados preliminares, possibilita propor uma ordem pseudo-cronológica do ciclo celular, identificando as amostras em sua fase no ciclo celular. Com isso, se obtêm informações mais detalhadas da evolução da célula no ciclo.

OBJETIVOS

- **Obtenção e preparação dos dados de medidas de expressão gênica de células única de *Mus Musculus de 288 amostras [2].***
- **Obtenção do transcriptograma e aplicação do PCA.**
- **Proposição da ordem pseudo-cronológica.**

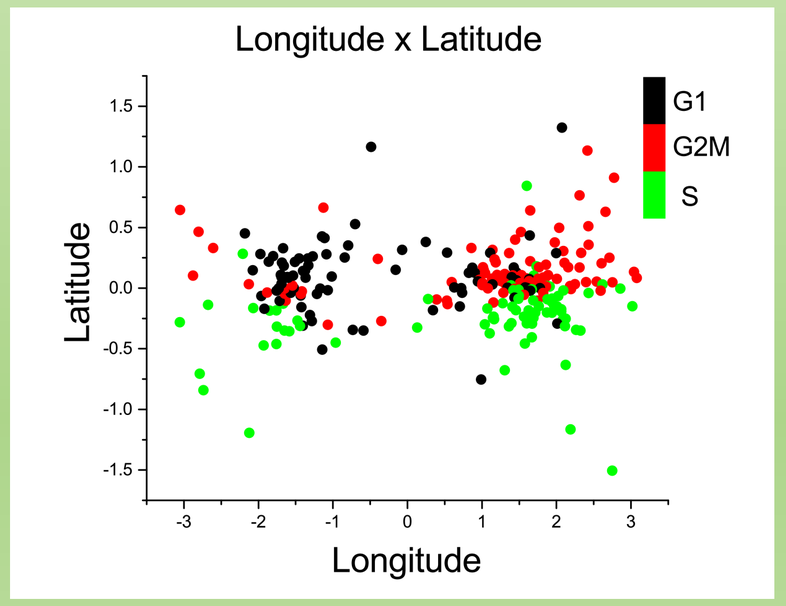
RESULTADOS

Realizando o método PCA para raio 30 em amostras normalizadas por RPKM, obtém-se a seguinte variância relativa acumulada:

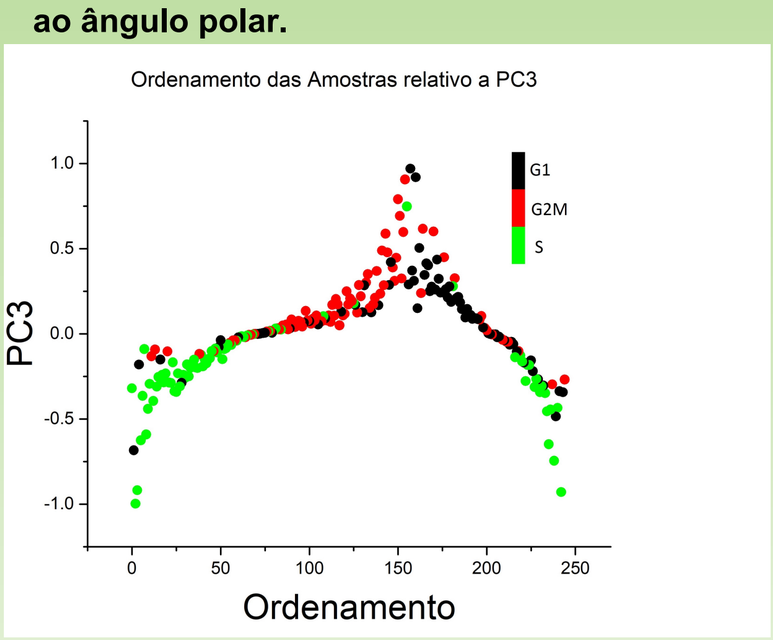


Com isso, observa-se que as 3 primeiras componentes principais apresenta em torno de 80% da informação dos dados.
Gerando um gráfico com as 3 principais componentes, normalizando à 1, obtém-se o seguinte:

Observa-se que a segunda componente (PC2), separa separa as amostras da fase G1 em relação às amostras em G2M e S. Já a terceira componente(PC3) , separa as amostras nas fase G1 e G2M em relação à fase S.
Fazendo uma transformação de coordenadas, passando o sistema inicial retangular em 3 dimensões para uma ângular análogo ao sistema geográfico, ficando mais evidente a separação das fases do ciclo celular.



Agora realizando uma nova mudança de coordenadas, para polares, fazendo a origem do novo sistema na origem do sistema anterior, pode-se ordenar as amostras relativa ao ângulo polar.



PRÓXIMO PASSO

- **Validação do ordenamento: Analisar grupos de genes, que estão associados a funções específicas na célula, caracterizando as amostras na respectiva fase do ciclo celular.**

[1]da Silva, S.R.M., Perrone, G.C., Dinis, J.M. and de Almeida, R.M.C. BMC Genomics, 15, 1181 (2014)
[2] Florian Buettner, Kedar N Natarajan, F Paolo Casale, Valentina Proserpio, Antonio Scialdone, Fabian J Theis, Sarah A Teichmann, John C Marioni & Oliver Stegle.http://www.nature.com/nbt/journal/v33/n2/full/nbt.3102.html Volume 33 No 2:155 - 160 (2015), PMID:25599176