

# Relatório Complementar à Tese de Mestrado (Tiago Lubiana Alves)

Tiago Lubiana Alves

8/21/2020

## Teor do documento

Este documento complementar à dissertação de mestrado visa atender aos requerimentos levantados pelo parecerista do relatório rejeitado pela FAPESP.

O parecer que será abordado é:

*"O relatório científico apresentado, principalmente no capítulo de metodologia, é bastante superficial. Esperavam-se maiores detalhes acerca das estratégias ali descritas e, também, das implementações de software discutidas ao longo do documento."*

O relato do orientador indica que o projeto foi executado com sucesso, havendo, inclusive, um manuscrito em produção. Desta maneira, é natural esperar que o relatório seja bastante detalhado."

Em específico, as seguintes observações:

*"A metodologia dos algoritmos implementados nos pacotes FCBF e fcoex, apresentada no Capítulo 3, é bastante superficial nos detalhamentos acerca dos passos: a) categorização da expressão gênica; b) identificação de módulos fcoex; c) análises de super-representação; d) identificação de módulos; e) análises de reagrupamento. Em particular, para cada um destes tópicos, um curto parágrafo é apresentado, sem informações de fato metodológicas acerca das estratégias. Tratando-se de um trabalho premiado em evento específico da área, espera-se que a dissertação contenha detalhamentos mais precisos acerca do entendimento, por parte do bolsista, nos tópicos apresentados."*

Dessa forma, foi refeito o Capítulo 3 da dissertação, tendo em vista atender o detalhamento metodológico desejado.

Como sugestão do parecerista, a ementa foi feita utilizando RMarkdown.

## Métodos - Revisado após parecer da FAPESP

Neste trabalho, almejamos desenvolver novas formas de analisar dados de células únicas transpondo ferramentas do campo do aprendizado de máquinas. Aqui descrevemos como obtivemos os conjuntos de dados públicos para teste e validação, como funcionam as ferramentas utilizadas para processamento desses dados e a detalhes metodológicos acerca dos algoritmos implementados nos pacotes *FCBF* e *fcoex*.

## Dados utilizados

O conjunto de dados pbmc3k versão 3.0.0 contém dados de 2700 observações ("células") mononucleares de sangue humano, foi obtido por meio do pacote *SeuratData* disponível em <https://github.com/satijalab/seurat-data>

O conjunto de dados SCP162 (de desenvolvimento do peixe-zebra) foi manualmente baixado do portal Broad Single Cell ([https://singlecell.broadinstitute.org/single/\\_cell/study/](https://singlecell.broadinstitute.org/single/_cell/study/)).

O conjunto de dados original contém 38732 observações (“células”), das quais foram mantidas as observações referentes ao estágio de 75% de epibolia (valor na coluna “Stage” igual a “08.0-75%”). O conjunto de dados usado para análise, então, conteve 6178 observações.