

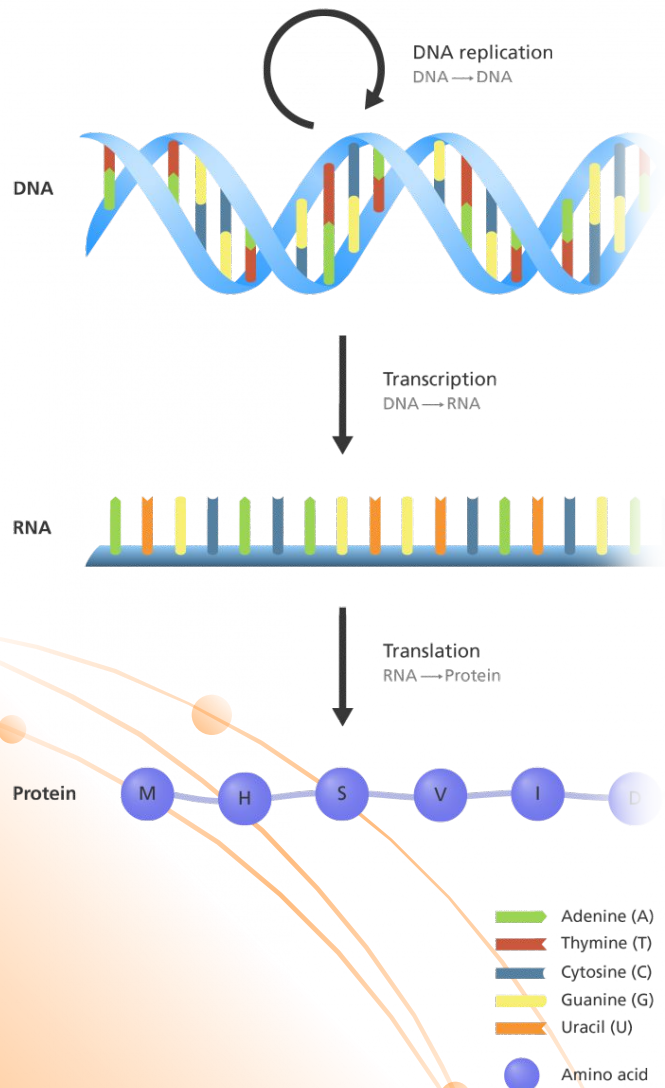
Avaliação do método *fcoex* em redes de coexpressão de genes

Projeto de Redes Complexas

Vinícius Werneck Salazar
Prof. Daniel Ratton Figueiredo
UFRJ - Novembro 2019



Introdução



Sequenciamento de RNA single cell

- Técnica de biologia molecular para quantificar a expressão de genes
- Usada para identificar tipos celulares (atribuir “labels”)

Redes de coexpressão de genes

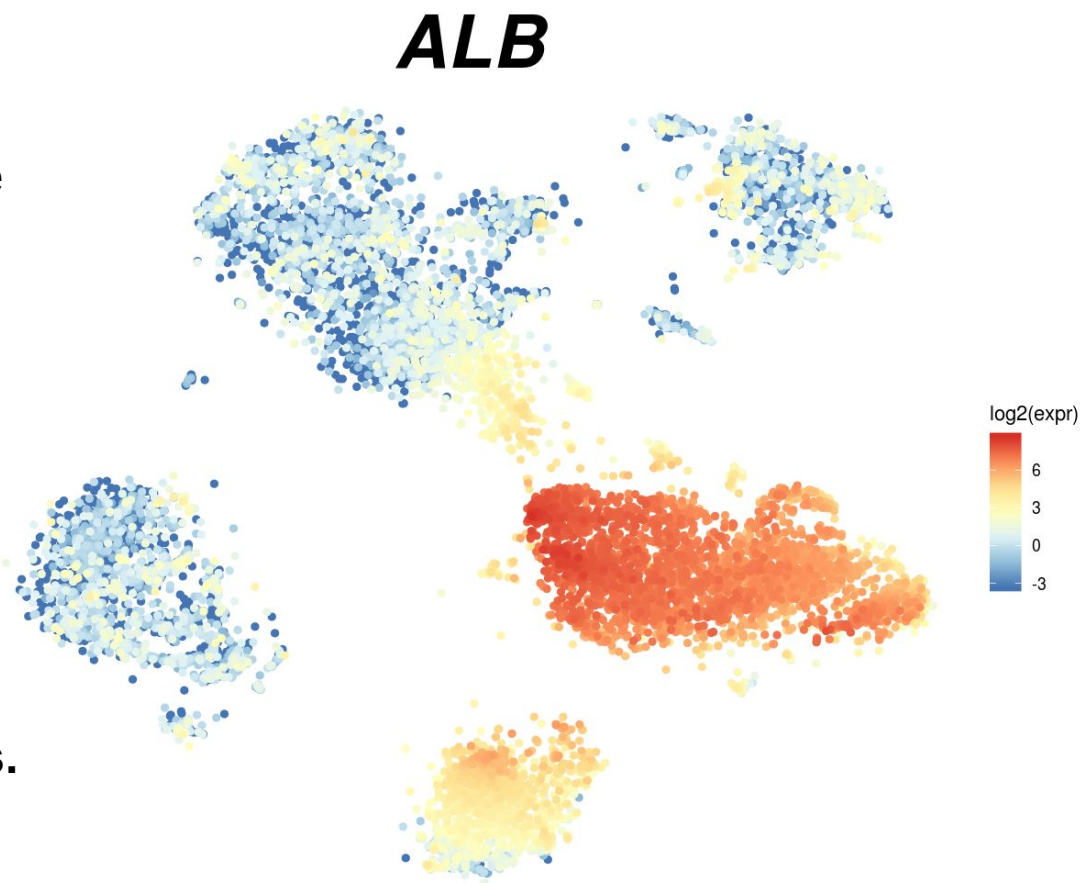
- Grafos não direcionados
- Cada gene é um vértice
- Se são expressos simultaneamente, são ligados por uma aresta



Redes de coexpressão de genes

Atlas celulares

- Representados através de redução de dimensionalidade (tSNE)*
- Ao lado, expressão do gene **ALB** em células hepáticas humanas
- Muito usadas em pesquisas biomédicas, no entanto **métodos para scRNA-seq ainda pouco desenvolvidos.**



Aizarani, N. *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* (2019).

* t-distributed stochastic neighbor embedding



O método *fcoex*

Redes de coexpressão de dados de scRNA-seq

- WGCNA → corr. Pearson → não indicado para scRNA-seq
- Pacotes de scRNA-seq não incluem módulos de coexpressão (e.g. Seurat, OSCA)
- Novo método, *fcoex*, implementado no Bioconductor
- Identificação de populações alternativas de células


[Home](#)
[Install](#)
[Help](#)

[Home](#) » [Bioconductor 3.10](#) » [Software Packages](#) » *fcoex*

fcoex

platforms [all](#) rank [1771 / 1823](#) posts [0](#) in Bioc [< 6 months](#)
 build [ok](#) updated [before release](#) dependencies [136](#)

DOI: [10.18129/B9.bioc.fcoex](#) [f](#) [t](#)

FCBF-based Co-Expression Networks for Single Cells

Bioconductor version: Release (3.10)

The *fcoex* package implements an easy-to use interface to co-expression analysis based on the FCBF (Fast Correlation-Based Filter) algorithm. It was implemented specifically to deal with single-cell data. The modules found can be used to redefine cell populations, unveil novel gene associations and predict gene function by guilt-by-association. The package structure is based on the CEMiTool package.

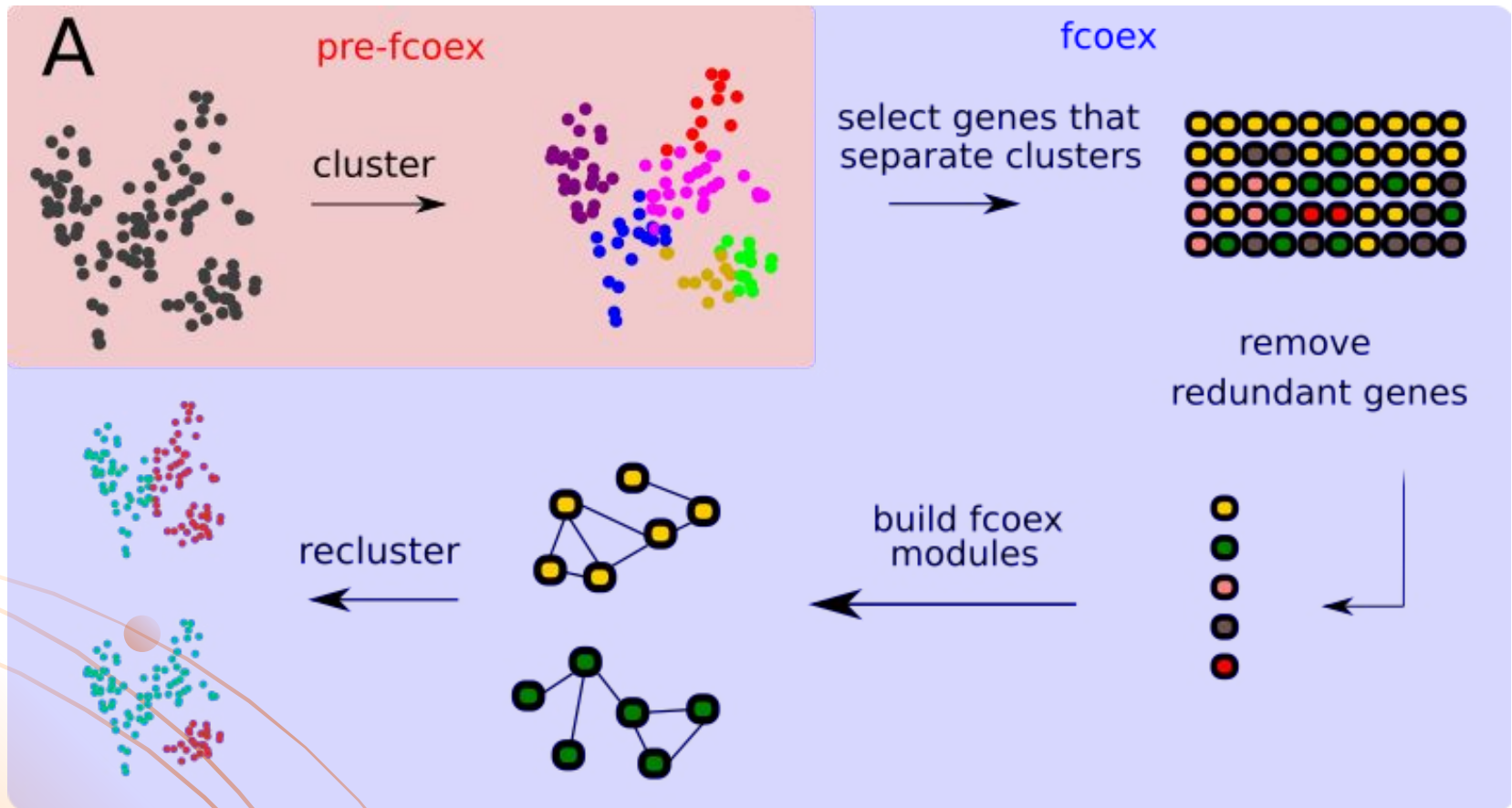
Author: Tiago Lubiana [aut, cre], Helder Nakaya [aut, ths]

Maintainer: Tiago Lubiana <tiago.lubiana.alves at usp.br>

Citation (from within R, enter `citation("fcoex")`):

Lubiana T, Nakaya H (2019). *fcoex: FCBF-based Co-Expression Networks for Single Cells*. R package version 1.0.0.

O método *fcoex*



Overview do pipeline do *fcoex*: Labels são previamente assinalados (via Seurat, por exemplo), genes com baixa correlação são removidos, e clusters são reconstruídos.



Método de correlação

fcoex é baseado no algoritmo FCBF, ou “fast correlation-based filter” (Yu & Liu, 2003)

- Conjunto de heurísticas para selecionar features de baixa redundância
- Assume um conjunto ergódico (probabilidade invariante)
- Correlação é baseada no conceito de ***Symmetrical Uncertainty***
- Ideia de ***over representation***

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b P(x_i)$$

Entropy

$$H(X|Y) = - \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(y_j)}$$

Conditional Entropy

$$I(X; Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x, y) \log \left(\frac{p(x, y)}{p(x) p(y)} \right),$$

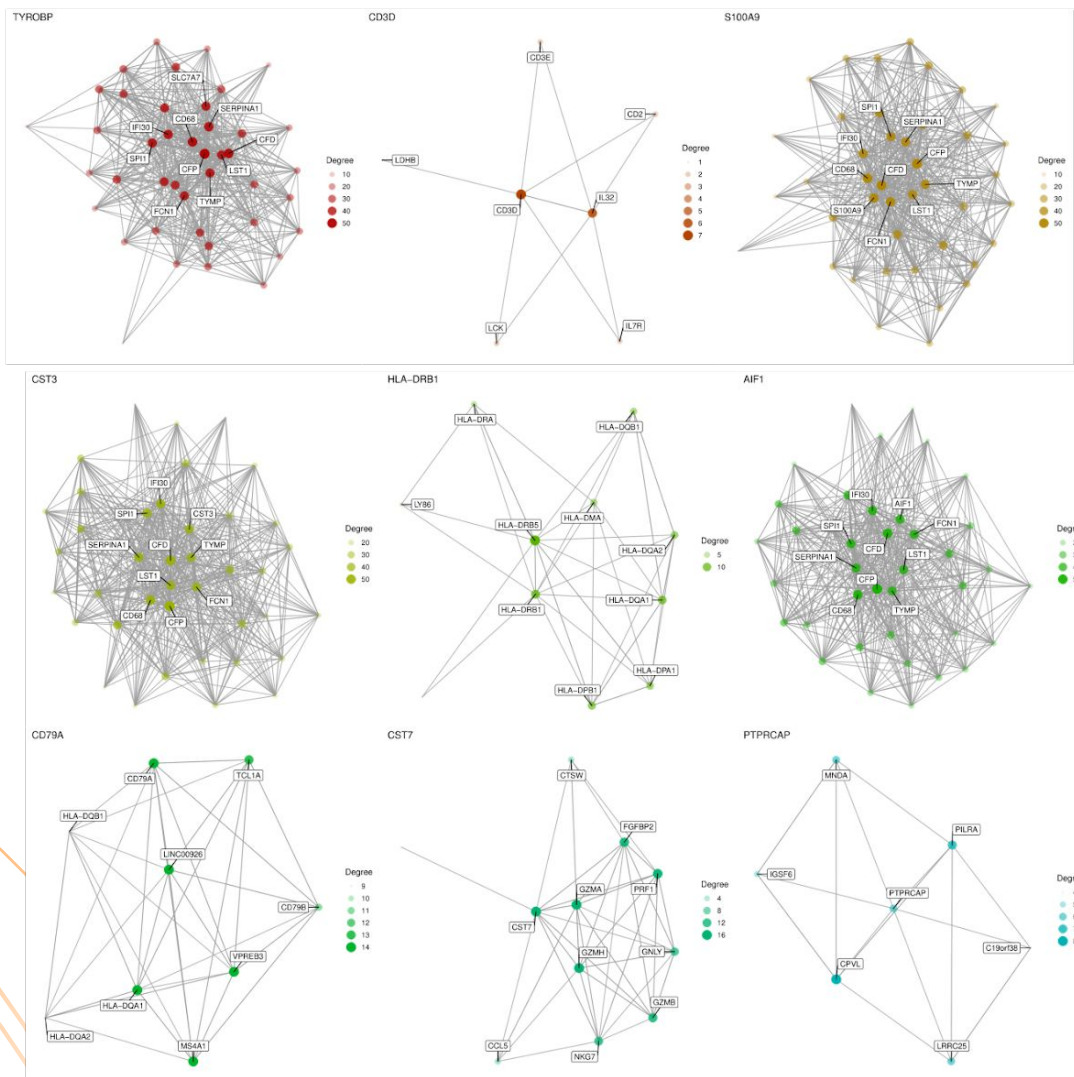
Information Gain

$$U(X, Y) = 2R = 2 \frac{I(X; Y)}{H(X) + H(Y)}$$

Symmetrical Uncertainty

Redes resultantes da aplicação do *fcoex*

A





Sobre redes de coexpressão e módulos

Considerações importantes:

1. *“Coexpression network biology has several open questions regarding formality.”
(...)*





Sobre redes de coexpressão e módulos

Considerações importantes:

1. *“Coexpression network biology has several open questions regarding formality.” (...)*
2. *“The meaning of different topological metrics is cloudy for coexpression networks, and to make ontological affirmations based on them would require extensive theoretical work.” (...)*





Sobre redes de coexpressão e módulos

Considerações importantes:

1. *“Coexpression network biology has several open questions regarding formality.” (...)*
2. *“The meaning of different topological metrics is cloudy for coexpression networks, and to make ontological affirmations based on them would require extensive theoretical work.” (...)*
3. *“However, the complex nature of these concerns can lead to ‘analysis paralysis’, if they are to be addressed before any application of coexpression networks.”*



Proposta de projeto

Objetivos:

- Analisar a estrutura de redes antes e depois da aplicação do método *fcoex*
- Replicar resultados como um *end user* do pacote do Bioconductor
- Entender como a aplicação do método afeta a estrutura das redes
- Discutir implicações teóricas para análise de redes de coexpressão, **especialmente no contexto de dados de scRNA-seq**
- Dessa forma, auxiliar na formalização do método *fcoex* e aprofundar a compreensão de como ele afeta as redes de coexpressão



Proposta de projeto

Objetivos:

- Analisar a estrutura de redes antes e depois da aplicação do método *fcoex*
- Replicar resultados como um *end user* do pacote do Bioconductor
- Entender como a aplicação do método afeta a estrutura das redes
- Discutir implicações teóricas para análise de redes de coexpressão, **especialmente no contexto de dados de scRNA-seq**
- Dessa forma, auxiliar na formalização do método *fcoex* e aprofundar a compreensão de como ele afeta as redes de coexpressão

O método “captura bem um fenômeno”, mas as implicações ainda são um pouco vagas.



Referências

1. Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* 4, 24–26 (2008).
2. Junker, B. H. & Schreiber, F. *Analysis of biological networks*. 2, (Wiley Online Library, 2008).
3. Lubiana, T. & Nakaya, H. fcoex: FCBF-based Co-Expression Networks for Single Cells. (2019). doi:10.18129/B9.bioc.fcoex
4. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008, P10008 (2008).
5. Yu, L. & Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. in *Proceedings of the 20th international conference on machine learning (ICML-03)* 856–863 (2003).
6. Aizarani, N. *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* (2019). doi:10.1038/s41586-019-1373-2