

Universidade Federal do ABC

Edital 04/2022 - PIC/PIBIC/IBITI/PIBIC-AFProjeto de Pesquisa

Transcriptômica em célula individual como ferramenta para estudo de doenças

neurodegenerativas: implicações para a doença de Alzheimer

Palavras chave: mRNA, genótipo, fenótipo, cérebro, sequenciamento de última geração

RESUMO

A doença de Alzheimer é uma doença neurodegenerativa que acomete milhões de pessoas. Essa patologia, objeto de inúmeros estudos científicos, progride causando a morte de células neuronais e atrofia de tecidos cerebrais e ainda não possui cura ou tratamento eficaz, tornando relevante a compreensão de aspectos biológicos e fatores agravantes do desenvolvimento dessa patologia. Portanto, serão abordadas neste projeto análises de bioinformática de transcriptomas de célula única através da plataforma de sequenciamento de última geração (NGS), utilizando dados de acesso público de amostras de cérebro post-mortem de pacientes com a doença de Alzheimer.

INTRODUÇÃO

Sequenciamento de RNA

Conhecer os elementos funcionais do genoma e informar sobre os constituintes moleculares de células e tecidos tem se revelado um recurso valioso para o entendimento de diferentes condições fisiológicas e fisiopatológicas, podendo ser alcançado pela identificação do conjunto completo de transcritos em uma célula, ou seja, o transcriptoma. O sequenciamento de RNA (RNA-Seq) é uma abordagem recentemente desenvolvida para o perfil do transcriptoma que usa tecnologias de sequenciamento de última geração (NGS) para revelar a presença e a quantidade de RNA em uma amostra biológica em condições específicas ou em um ponto de tempo específico. É um método de sequenciamento de DNA de alto rendimento que apresenta vantagens claras sobre as abordagens existentes previamente (Wang et al., 2009). Em geral, a população de RNA total é convertida em uma biblioteca de fragmentos de DNA complementar (cDNA) com adaptadores ligados a uma ou ambas as extremidades. Cada molécula é então sequenciada de uma maneira de alto rendimento para obter sequências curtas de uma extremidade (sequenciamento de extremidade única) ou de ambas as extremidades (sequenciamento de extremidade dupla). Após o sequenciamento, as leituras resultantes são alinhadas a um transcrito de referência ou genoma de referência para produzir um mapa de transcrição em escala genômica que consiste em estrutura transcricional e/ou nível de expressão de cada gene. O RNA-Seq é usado para analisar o transcriptoma celular em constante mudança. Especificamente, permite observar transcrições de genes alternativos, modificações pós-transcricionais, fusão gênica, mutações/SNPs e mudanças na expressão gênica ao longo do tempo, ou diferenças na expressão gênica em diferentes grupos ou tratamentos (Wang et al., 2009).

De fato, a análise do transcriptoma apareceu como uma estratégia poderosa para resolver os desafios de mapeamento de genótipos para fenótipos em biologia e medicina. No entanto, embora todas as células do nosso corpo compartilhem genótipos idênticos, a informação do transcriptoma em cada célula reflete a atividade de apenas um subconjunto de genes. Além disso, o transcriptoma dos diversos tipos de células é único e o sequenciamento convencional da população em massa é limitado, pois fornece apenas o sinal de expressão médio para uma coleção de células (Hwang et al., 2018). Mesmo em tipos de células semelhantes, a expressão gênica pode ser heterogênea, como sugerem evidências recentes (Huang, 2009, Hwang et al., 2018, Shalek et al., 2014). Nesse sentido, a variabilidade célula a célula está ausente em estudos que usam sequenciamento em massa e baseiam-se na suposição de que as células de um determinado tecido são homogêneas (Figura 1). Assim, uma compreensão precisa do transcriptoma em células individuais é crucial para entender como a expressão gênica pode promover condições benéficas ou prejudiciais e, assim, esclarecer seu papel nas funções celulares (Hwang et al., 2018).

O sequenciamento de um transcriptoma inteiro ao nível de uma única célula data do início dos anos 90. Para isso os transcritos da célula (RNA mensageiro - mRNA) são convertidos em cDNA, e amplificados por PCR (Brady e Iscove, 1993, Eberwine et al., 1992). O cDNA é então fragmentado, marcado com adaptadores específicos reconhecidos pela plataforma de NGS para que possam ser sequenciados. A plataforma de NGS para análise de transcriptoma de célula única foi descrita pela primeira vez em 2009 (Tang et al., 2009) e, desde então, tem havido um interesse crescente em adquirir visualizações de alta resolução da heterogeneidade de célula única em escala global (Hwang et al., 2018).

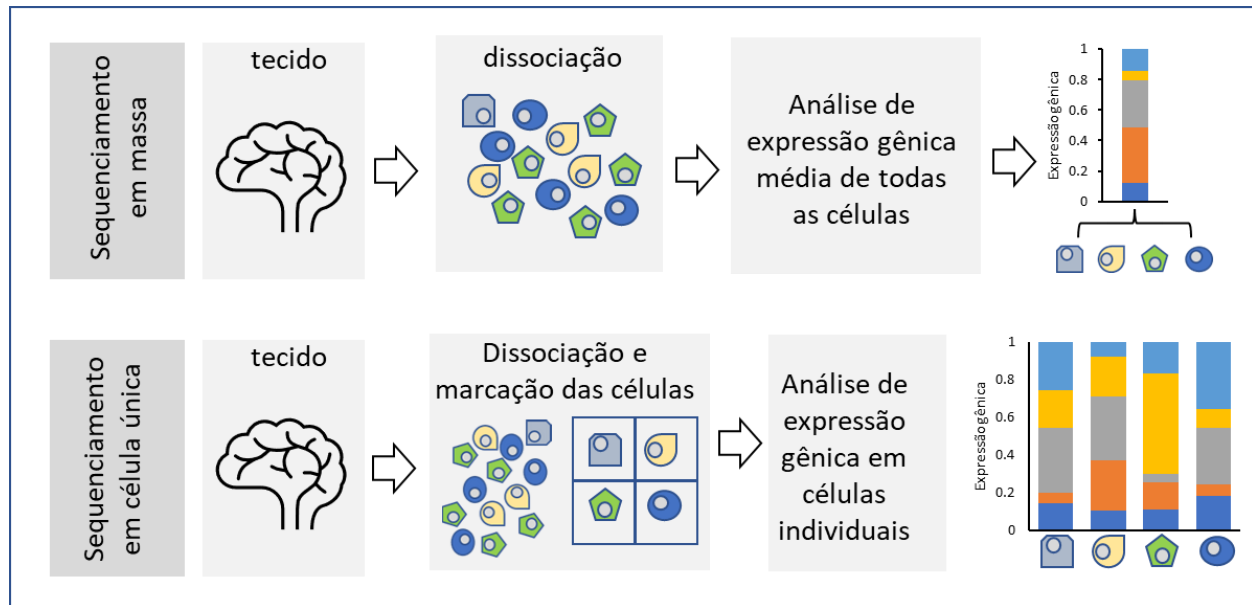


Figura 1. Esquema simplificado das técnicas para sequenciamento de RNA, ressaltando diferenças entre sequenciamento em massa e sequenciamento em célula individual. O resultado do sequenciamento em massa é a expressão gênica média de todas células presentes no tecido. Já no sequenciamento em célula individual é possível estudar a expressão gênica em cada célula e assim detectar heterogeneidades na expressão gênica condicionadas ao tipo celular.

RNA-Seq na doença de Alzheimer

A doença de Alzheimer é uma patologia neurodegenerativa, com alta prevalência no mundo para a qual até hoje não existe cura ou tratamento eficaz. Essa patologia é caracterizada principalmente pelo acúmulo de agregados proteicos disfuncionais (Beta amiloide e proteína Tau) que leva a morte dos neurônios que resulta em atrofia no tecido cerebral. Por se tratar de uma doença multifatorial, não existe ainda uma compreensão extensiva dos fatores que levam ao desenvolvimento da patologia, mas sabe-se que a morte neuronal é o resultado de um longo processo que envolve a participação de diferentes tipos celulares encefálicos como astrócitos, microglia, oligodendrócitos e vasculatura cerebral. Desta maneira, conhecer as particularidades

das respostas celulares nos encéfalos dos pacientes, traz grande potencial para melhor compreensão da patologia e embasamento para estratégias terapêuticas.

A fisiopatologia do cérebro de pacientes com DA, bem como as alterações de expressão global reveladas por análises de transcriptoma em massa são amplamente estudadas. Os níveis gerais de RNA do tecido cerebral com DA foram descritos anteriormente, mas esses estudos não distinguem entre os tipos de células, o que pode mascarar as alterações que ocorrem em tipos de células menos abundantes. A oportunidade de realizar análise de transcriptoma de célula única abre novas perspectivas para a compreensão da DA. Um dos primeiros estudos que explorou a tecnologia de sequenciamento de RNA em célula individual foi publicado por Mathys et al. (Mathys et al., 2019). Neste estudo, os pesquisadores utilizaram amostras de cérebro de córtex pré-frontal post-mortem de pacientes com altos níveis de patologia de DA e controles pareados por idade. Eles mostraram que, em pessoas com DA, os genes relacionados à mielinização foram afetados tanto nos neurônios quanto nos oligodendrócitos, as células que produzem mielina, em estágios iniciais da doença. Em estágios posteriores, a maioria dos tipos de células tinha padrões muito semelhantes de mudança de expressão gênica, revelando genes relacionados à resposta ao estresse, morte celular programada e maquinaria celular necessária para manter a integridade da proteína. Também foi descrito que as células cerebrais de homens e mulheres variam significativamente na forma como seus genes respondem à doença. Estudos como este (Mathys et al., 2019) permitem a identificação de potenciais alvos para o tratamento da doença, bem como o desenvolvimento de modelos como por exemplo utilizando células-tronco pluripotentes induzidas, ou animais transgênicos para aprofundar o estudo de algumas das principais vias celulares que são identificadas como associadas à DA. Desde a publicação do estudo de Mathys, em 2019, outros estudos focados em entender as alterações transcriptômicas em amostras post-mortem de pacientes

com DA foram publicados (Olah et al, 2020, Leng et al, 2021, Grubman et al, 2019, Lau et al, 2020, Zhou et al, 2020,). Não há dúvidas de que os estudos e respectivas análises trazem conhecimento de imensurável importância para o melhor entendimento desta patologia, que hoje afeta mais de 50 milhões de pessoas no mundo. Mas além disso, o depósito dos dados de sequenciamento em databases, e acesso público a esses dados, permitem explorações adicionais e outros focos de análise que podem também colaborar para o desenvolvimento da área.

Neste projeto, iremos explorar dados de acesso público de transcriptômica de amostras de cérebro post-mortem de pacientes diagnosticados com DA, com o objetivo de capacitar o aluno para realizar análises de bioinformática de sequenciamento de RNA em célula única.

OBJETIVOS

- Treinamento em software R e RStudio.
- Treinamento para uso do pacote Seurat para análise de sequenciamento de RNA em célula individual
- Análise dos dados de acesso público de sequenciamento de célula individual em amostras de pacientes com DA.

MATERIAL E MÉTODOS

- **Treinamento em Software R e RStudio**

O treinamento será feito pela orientadora com auxílio de cursos online gratuitos para aprender habilidades básicas de linguagem R e ferramenta RStudio.

Cursos de apoio online:

- ✓ DidáticaTech: Curso de R online para iniciantes

<https://didatica.tech/curso-de-r-online-para-iniciantes/>

- ✓ Curso R com RStudio (21 aulas, canal escola de inteligência artificial – youtube):

<https://www.youtube.com/channel/UC6w56wRjrbbZk-cX46SuxPA/videos>

- ✓ **Treinamento para uso do pacote Seurat para análise de sequenciamento de RNA**

O pacote Seurat, desenvolvido pelo laboratório do pesquisador Rahul Satija (<https://as.nyu.edu/content/nyu-as/as/faculty/rahul-satija.html>), é amplamente utilizado para análise de RNA-seq em célula individual. Para treinamento do uso do pacote Seurat serão seguidas as Vignettes introdutórias disponibilizadas no site dos desenvolvedores do pacote:

https://satijalab.org/seurat/articles/get_started.html

A Vignette “**Guided tutorial**” (https://satijalab.org/seurat/articles/pbmc3k_tutorial.html) será realizada com os dados públicos dos dados de transcriptômica de células periféricas sanguíneas mononucleares (PBMC) disponibilizados pela empresa 10X Genomics:

<https://www.10xgenomics.com/resources/datasets?query=&page=1&configure%5Bfacets%5D%5B0%5D=chemistryVersionAndThroughput&configure%5Bfacets%5D%5B1%5D=pipeline.version&configure%5BhitsPerPage%5D=500>

As Vignettes “**Introduction to scRNA-seq integration**”, “**Data visualization methods in Seurat**”, “**Differential expression testing**”, “**Dimensional Reduction**”, serão executadas com os dados disponíveis no pacote “*SeuratData*”:

<https://github.com/satijalab/seurat-data>

- **Análise dos dados de acesso público de sequenciamento de célula individual em amostras de pacientes com DA.**

Os dados públicos de sequenciamento do manuscrito

<https://doi.org/10.7303/syn21788402> (Leng et al, 2021) serão baixados da plataforma Gene Expression Omnibus (GEO) - accession code GSE147528 (dados não processados) ou da plataforma synapse.org - Synapse id: syn21788402 (dados processados). Os dados serão re-analizados utilizando as habilidades adquiridas com os treinamentos em RStudio e Seurat.

CRONOGRAMA

	1° quadrimestre	2° quadrimestre	3° quadrimestre
Treinamento em R e RStudio	✓		

Treinamento Seurat		✓	
Análise de dados de acesso público			✓
Redação relatórios parcial e final		✓	✓

REFERÊNCIAS

- Brady, G. & Iscove, N. N. 1993. Construction of cDNA libraries from single cells. *Methods Enzymol*, 225, 611-23.
- Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. & Coleman, P. 1992. Analysis of gene expression in single live neurons. *Proc Natl Acad Sci U S A*, 89, 3010-4.
- Grubman, A., Chew, G., Ouyang, J.F. et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer’s disease reveals cell-type-specific gene expression regulation. *Nat Neurosci* 22, 2087–2097 (2019).
- Huang, S. 2009. Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136, 3853-62.
- Hwang, B., Lee, J. H. & Bang, D. 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med*, 50, 96.
- Kenigsbuch M, Bost P, Halevi S, Chang Y, Chen S, Ma Q, Hajbi R, Schwikowski B, Bodenmiller B, Fu H, Schwartz M, Amit I. A shared disease-associated oligodendrocyte

signature among multiple CNS pathologies. *Nat Neurosci.* 2022 Jun 27. doi:
10.1038/s41593-022-01104-7. Online ahead of print.

Leng, K., Li, E., Eser, R. et al. Molecular characterization of selectively vulnerable neurons in
Alzheimer's disease. *Nat Neurosci* 24, 276–287 (2021).

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M.,
He, L., Abdurrob, F., Jiang, X., Martorell, A. J., Ransohoff, R. M., Hafler, B. P., Bennett, D.
A., Kellis, M. & Tsai, L. H. 2019. Single-cell transcriptomic analysis of Alzheimer's disease.
Nature, 570, 332-337

Olah, M., Menon, V., Habib, N. et al. Single cell RNA sequencing of human microglia uncovers
a subset associated with Alzheimer's disease. *Nat Commun* 11, 6129 (2020).

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., Chen, P., Gertner, R. S.,
Gaublomme, J. T., Yosef, N., Schwartz, S., Fowler, B., Weaver, S., Wang, J., Wang, X.,
Ding, R., et al. 2014. Single-cell RNA-seq reveals dynamic paracrine control of cellular
variation. *Nature*, 510, 363-9.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch,
B. B., Siddiqui, A., Lao, K. & Surani, M. A. 2009. mRNA-Seq whole-transcriptome analysis
of a single cell. *Nat Methods*, 6, 377-82.

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics.
Nat Rev Genet, 10, 57-63.

Zhou, Y., Song, W.M., Andhey, P.S. et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med* 26, 131–142 (2020).