

## **ANÁLISE FUNCIONAL DO MICROBIOMA INTESTINAL DE CAMUNDONGOS DURANTE O DESENVOLVIMENTO DA CAQUEXIA INDUZIDA POR TRANSPLANTE DE CÉLULAS DE CÂNCER PULMONAR (LLC)**

Rafael dos Santos Gonçalves<sup>1</sup>; Regina Costa de Oliveira<sup>2</sup>, Fabiano B. Menegidio<sup>3</sup> Luiz Roberto Nunes<sup>4</sup>

1. Estudante do curso de Sistemas de Informação; rafaelga7@gmail.com
2. Coordenadora Programa de Pós-Graduação em Biotecnologia; reginaco@umc.br
3. Professor da Universidade de Mogi das Cruzes; fabiano.menegidio@biology.com.br
4. Orientador do Programa de Pós-Graduação em Biotecnologia da Universidade de Mogi das Cruzes; nunes1212@gmail.com

Área de Conhecimento: **Genética Molecular, Microbiologia.**

**Palavras-Chaves:** Caquexia; Microbioma; Micobioma

### **INTRODUÇÃO**

A caquexia é reconhecida como uma síndrome metabólica associada a diversas doenças subjacentes, como câncer, doença renal crônica, entre outras enfermidades. É caracterizada pela redução da massa muscular, diminuição dos estoques de gordura corpórea e inflamação crônica generalizada (EVANS *et al.*, 2008; FEARON *et al.*, 2011). Estudos recentes apresentam que alterações na composição da microbiota intestinal podem estar relacionadas ao desenvolvimento de doenças metabólicas e síndromes, como diabetes e obesidade e, desta forma, situações de desequilíbrio da microbiota denominadas de disbioses poderiam contribuir para o agravamento de infecções por patógenos, mas também para o estabelecimento de quadros inflamatórios crônicos, pois as alterações na microbiota interagem diretamente com o sistema imune do hospedeiro (ALMEIDA *et al.*, 2003; EVANS *et al.*, 2008; FEARON *et al.*, 2011; ZHANG *et al.*, 2015). O RNA ribossomal 16S rRNA, que possui 9 regiões hipervariáveis intercaladas com regiões altamente conservadas (FORDE *et al.*, 2013), é utilizado para identificar as modificações e categorizar a microbiota intestinal associadas a patologias. Sendo a caquexia uma síndrome relacionada a disfunções no metabolismo energético, notam-se que alterações na composição da microbiota também estejam correlacionadas com o desenvolvimento de quadros caquéticos, como constatado nos casos de obesidade e diabetes (BINDELS *et al.*, 2015, 2016). Conhecendo-se, ao mesmo tempo, a microbiota e o microbioma de uma comunidade de microrganismos, é possível recorrer a bancos de dados que possam identificar os taxa que compõem a população estudada e reconstruir as principais vias metabólicas e biossintetizadores dos microrganismos ali presentes, sendo possível fazer previsões a respeito do potencial metabólico da comunidade (ou seja, inferir os principais tipos de metabólitos que podem ser por ela produzidos). Neste caso, seria possível inferir, também, as principais alterações bioquímicas e metabólicas que ocorrem em uma comunidade microbiana, quando um quadro de disbiose se estabelece.

### **OBJETIVOS**

O presente trabalho teve como objetivo realizar uma comparação funcional dos microbiomas murinos no período de pré e pós caquexia, reconstruindo as principais vias metabólicas e biossintetizadores dos microrganismos ali presentes, com a utilização de ferramentas de bioinformática para permitir a identificação das alterações metabólicas que podem ocorrer na comunidade de microrganismos, durante o desenvolvimento dessa síndrome.

## METODOLOGIA

A análise funcional conduzida neste projeto utilizou dados de sequenciamento NGS da região hipervariável V4-V5 de amostras de fezes de camundongos, obtidas de experimentos decorrentes do desenvolvimento do Projeto FAPESP nº 2017/08112-3. Estes dados foram submetidos a um fluxo de trabalho desenvolvido no próprio laboratório, sendo ele dividido nas seguintes etapas: análise de qualidade, pré-processamento, processamento e pós-processamento. Durante a etapa de análise de qualidade, as ferramentas FASTQC (ANDREWS, 2010) e MULTIQC (EWELS *et al.*, 2016) foram utilizadas com o propósito de analisar a qualidade de todas as sequências provenientes de todas as bibliotecas obtidas após o sequenciamento. Ao decorrer das etapas de pré-processamento e processamento, o pacote de ferramentas QIIME (CAPORASO *et al.*, 2010) foi utilizado. O foco da etapa de pré-processamento foi realizar uma limpeza de qualidade nos dados, sendo assim, o *script multiple\_join\_paired\_ends.py* foi utilizado com a finalidade de realizar a junção entre todas as bibliotecas *forward* e *reverse*. Após a junção, o processo de *Splitting Libraries* foi realizado, sendo o mesmo responsável por renomear o cabeçalho das sequências, fazer o corte de qualidade, concatenar todas as sequências e converter o arquivo para a extensão FASTA. Antes de finalizar esta etapa, os *scripts identify\_chimeric\_seqs.py* e *filter\_fasta.py* foram utilizados para a remoção de todas as sequências que foram identificadas como quimeras. Após a filtragem das quimeras, deu-se início a etapa de processamento, sendo ela responsável pela criação da *OTU table* (unidades taxonômicas operacionais) e criação do microbioma central (core). O *script pick\_open\_reference\_otus.py* foi responsável pela criação da *OTU table* e é dividido em quatro fases: a clusterização (agrupamento de sequências por nível de similaridade), atribuição taxonômica dos grupos criados na fase de clusterização, alinhamento das sequências e criação da árvore filogenética, sendo que nas três primeiras fases o banco de dados do *Greengenes* (DESANTIS, 2006) foi utilizado como referência. Após sua criação, a *OTU table* foi submetida a *scripts* de filtragem do pacote QIIME que possuem os seguintes objetivos: remoção de mitocôndrias e cloroplastos contaminantes, descarte de *singletons* e filtragem de OTUs por número de observações. O último *script* utilizado na etapa de processamento foi o *compute\_core\_microbiome.py*, sendo ele responsável por criar uma nova *OTU table* contendo apenas as OTUs presentes em 80% de todas as amostras. Na última etapa do fluxo de trabalho, pós-processamento, foram utilizados dois *softwares*, o PICRUSt (LANGILLE *et al.*, 2013) e a ferramenta web MicrobiomeAnalyst (DHARIWAL *et al.*, 2017). O algoritmo utilizado no PICRUSt é implementado em várias etapas, seu objetivo principal é realizar a predição funcional de todo o metagenoma utilizando o arquivo gerado pelo método core e um banco de dados de grupos de genes ortólogos (KEGG), resultando e uma tabela de contagens funcionais. O MicrobiomeAnalyst (DHARIWAL *et al.*, 2017) integra diferentes ferramentas estatísticas e técnicas de visualização para permitir que o usuário realize análises abrangentes a partir do estudo de microbiomas. O arquivo de contagens funcionais foi submetido a ferramenta web e foi realizado o teste LEfSe, sendo este teste um algoritmo para a descoberta e explicação de biomarcadores de alta dimensionalidade que identifica as características genômicas (genes, vias metabólicas e taxa) caracterizando as diferenças entre duas ou mais condições biológicas.

## RESULTADOS E DISCUSSÃO

O sequenciamento de todas as bibliotecas provenientes das fezes dos camundongos nos rendeu um valor absoluto de 50.287.754 leituras. Após a filtragem de quimeras, última fase da etapa de pré-processamento, o valor absoluto de sequências encontradas foi de 11.316.918. Nota-se uma grande queda no valor final da etapa de pré-processamento quando comparada ao valor inicial, sendo que um dos principais motivos é que há o processo de alinhamento das bibliotecas *forward* e *reverse*, fazendo com que as sequências das bibliotecas não sejam contadas de forma individual. Na etapa de processamento e após a criação e todos os processos de filtragens das *OTUs tables*, os seguintes valores foram

encontrados: *OTU table* com filtro de abundância de 0,001% e core em 80% de prevalência, o valor de observações foi de 269 e 3.805.553 de sequências. Esse resultado do core apresentou a separação dos três grupos analisados (grupo caquético, tumor-bearing e controle) ambos no período de 28 dias de coleta fecal. Para validar estes dados, diversos testes foram realizados, sendo um deles foi o teste aleatório balanceado, no qual três grupos foram criados e nomeados de caq-28, tb-28 e controle-28, sendo que cada grupo foi definido com quantidades de amostras iguais e aleatórias de grupos com diferentes classificações de caquexia, sendo que foi possível verificar que em todos os testes aleatórios balanceados não houve a separação dos grupos. Após a execução do PICRUSt (LANGILLE *et al.*, 2013) utilizando a *OTU table* gerada pelo core, uma nova tabela de contagens funcionais foi gerada e submetida ao MicrobiomeAnalyst (DHARIWAL *et al.*, 2017) para que o teste LEfSe fosse realizado. Através do teste foi possível verificar que dos 15 elementos funcionais mais representativos de todos os dados do core, 11 elementos estão presentes no grupo caquético e apenas 4 no grupo de camundongos que passaram pelo processo de inoculação do câncer porém não desenvolveram o câncer.

## CONCLUSÕES

As ferramentas utilizadas durante todas as etapas do fluxo de trabalho apresentaram bons resultados durante e após toda a sua execução, nos permitindo afirmar que a pipeline é apta para trabalhar com dados de microbioma. Sendo que ao decorrer da execução da *pipeline*, diversos testes com valores de parâmetros foram realizados para obter a maior qualidade e quantidade de dados possível para se trabalhar. Notou-se um enorme aumento nas pontuações de qualidade de todas as bibliotecas durante a etapa de pré-processamento, quando comparadas a análise de qualidade inicial dos dados. O método de criação de *OTU table* apenas contendo o microbioma central presente em, no mínimo, 80% das amostras analisadas, o core, apresentou bons resultados referente a separação dos grupos de camundongos e resolvendo o viés estatístico, no qual *OTUs* representativas em valores de abundância em apenas uma ou em poucas amostras foram excluídas. O software PICRUSt (LANGILLE *et al.*, 2013) realizou as predições funcionais de todo o microbioma analisado com base no arquivo gerado pelo método do core com toda a taxonomia definida e com um banco de dados de genes ortólogos (KEGG), fornecendo uma tabela de contagem funcional. Esta tabela foi submetida ao MicrobiomeAnalyst (DHARIWAL *et al.*, 2017) para realizar o teste estatístico LEfSe e analisar todos os elementos funcionais encontrados, sendo que dos 15 elementos funcionais mais representativos, 11 estão presentes no grupo caquético e apenas 4 no grupo de camundongos que passaram pelo processo de inoculação, porém não desenvolveram a caquexia. Sendo assim, objetivo do projeto foi concluído, sendo a próxima etapa realizar análises taxonômicas e funcionais mais a fundo dos dados obtidos para averiguar se apresentam ou não correlação coma síndrome metabólica caquética.

## REFERÊNCIAS

- ALMEIDA, P. *et al.* Efeitos de antimicrobianos, prebióticos, probióticos e extratos vegetais sobre a microbiota intestinal. **R. Bras. Zootec.** v.35, n.6, p.2359-2367, 2003.
- BINDELS, L.B., *et al.* Non digestible oligosaccharides modulate the gut microbiota to control the development of leukemia and associated aachexia in mice. **Plos One**, v.10, n.6, p.e0131009, 2015.
- ANDREWS, S. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Acesso 11 janeiro de 2019. 2010.
- Bindels, L.B., et al. Synbiotic approach restores intestinal homeostasis and prolongs survival in leukaemic mice with cachexia. **ISME J**, v.10, n.6, p.1456-1470, 2016.

CAPORASO, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. **Proc Natl Acad Sci USA**, v.108, n.1, p.4516-22, 2010.

DESANTIS, T. Z. et al. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microb* 72(7): 5069-5072.

DHARIWAL et al. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. **Nucleic acids research**, 45(W1), W180-W188.

EVANS, W.J., et al. Cachexia: a new definition. **Clin Nutr.**, v.27, n.6, p.793-799, 2008.

EWELS, P. et al. MultiQC: summarize analysis results for multiple tools and samples in a single report, **Bioinformatics**, Volume 32, Issue 19, 1 October 2016, Pages 3047–3048

Fearon, K, et al. Definition and classification of cancer cachexia: an international consensus. **Lancet Oncol.**, v.12, n.5, p.489-495, 2011.

FORDE, B.M. TOOLE. Next-generation sequencing technologies and their impact on microbial genomics. **Brief Funct Genomics**. V.12, p. 353-440, 2013.

LANGILLE, M. G. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. **Nat Biotechnol.**, v.31, n.9, p.814-21, 2013.

Zhang, Y., et al. Metagenomics: A New Way to Illustrate the Crosstalk between Infectious Diseases and Host Microbiome. **Int J Mol Sci.**, v.16, n.11, p.2; 2015.