DESENVOLVIMENTO DE UM BANCO DE DADOS DE TRANSCRITOS PARA LÚPUS ERITEMATOSO SISTÊMICO

Fernanda R. Paulino (PROVIC), Adriano Ferrasa (Orientador), e-mail: ferrasa@uepg.br

Universidade Estadual de Ponta Grossa / Departamento de Informática

Ciência da Computação (1.03.00.00-7) / Metodologia e Técnicas da Computação (1.03.03.00-6)

Palavras-Chave: *Análise de transciptoma, Lúpus Eritematoso Sistêmico, Banco de dados biológico.*

Resumo

O Lúpus Eritematoso Sistêmico (LES) é uma doença autoimune crônica que afeta múltiplos órgãos e sistemas, como pele, rins e sistema nervoso. Este trabalho desenvolveu um banco de dados biológico contendo dados de expressão gênica de amostras de LES. Para isso foi utilizando dados públicos de transcriptoma, e aplicada técnicas de bioinformática e *pipelines* computacionais para a análise e identificação de genes diferencialmente expressos entre amostras controle e de LES. Com isso buscou-se e criação de um recurso computacional que possa ser utilizado para pesquisas que envolvam o entendimento de mecanismos moleculares do LES.

Introdução

O transcriptoma é o conjunto completo de transcritos de um organismo, sendo composto de RNAs mensageiros, transportadores dentre outros (FIGUEIREDO, 2020). Através de dados de expressão de transcritos, pode-se observar padrões de expressão gênica de moléculas de RNA. No contexto do LES, uma doença autoimune e de origem poligênica, a análise destes padrões de expressão, pode fornecer meios para um maior entendimento sobre os mecanismos moleculares da doença. Através da criação de um banco de dados biológico com estas informações de expressão gênica de LES, buscou-se fornecer recursos para auxiliar pesquisas sobre a etiologia do LES.

Materiais e Métodos

Os dados de transcriptoma utilizados na análise, foram obtidos através do banco de dados público BioProject, disponibilizado pelo *National Center for Biotechnology Information* (NCBI), utilizando parâmetros específicos para busca de dados de expressão gênica de LES. Para a análise destes dados, foi desenvolvido um *pipeline* computacional que inclui etapas de filtragem de qualidade, mapeamento dos transcritos contra o genoma de referência humano (versão hg38), contagem e normalização de *reads* e análise estatística utilizando o pacote DESeq2 (LOVE; HUBER; ANDERS, 2014) e o software R. Os dados da análise foram anotados com metadados e organizados em formato de arquivo tabular.

Resultados e Discussão

Foram selecionados todos os estudos de transcriptoma público de LES, totalizando 1188 amostras. Com a aplicação do *pipeline* computacional foi possível a análise automatizada dos dados, resultando na identificação de genes diferencialmente expressos entre amostras controle e de LES. Os resultados desta análise expõem genes com expressão alterada em amostras LES, comparadas com amostras controle, os quais, através de estudos específicos, podem contribuir para um entendimento mais aprofundado sobre o desenvolvimento desta doença. Todas as informações desta análise foram disponibilizadas através da criação de um banco de dados biológico o qual poderá ser disponibilizado para o acesso da comunidade científica.

Conclusão ou Conclusões

O *pipeline* computacional desenvolvido oferece um processo eficiente para a análise automatizada de dados de expressão gênica, independentemente do tipo de amostras. Ainda, a criação de um banco de dados de transcriptoma para LES fornece meios para pesquisas que visam compreender os mecanismos moleculares de LES.

Agradecimentos

Ao Prof. Dr. Luciano Jose Senger (LCAD, UEPG), Prof. Dr. Alysson Muotri (UCSD, USA) e Prof. Dr. Roberto H. Herai (LaBiN, PUCPR), pela colaboração na disponibilização de estrutura computacional para este projeto.

FIGUEIREDO, J. E. F.; VASCONCELOS, M. J. V. **Função dos miRNAs na regulação gênica pós transcricional e perspectivas para o melhoramento de plantas**. Sete Lagoas: Embrapa Milho e Sorgo, 2020.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 1–34, 2014.

ISSN 1676-0093