UNIVERSIDADE FEDERAL DE PERNAMBUCO

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CENTRO DE INFORMÁTICA

2014.1

Análise de Abordagens para o Problema de Detecção de Copy Number Variations

---

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Lucas Menezes Veras (lmv@cin.ufpe.br)

Orientadora: Katia Silva Guimarães (katiag@cin.ufpe.br)

Recife, maio de 2014

Sumário

Contexto	3
Objetivo	4
Cronograma	4
Referências	5
ASSINATURAS	6

Contexto

O estudo de variações genéticas é de grande relevância para diversas áreas da biologia, e tem sido favorecido pelo desenvolvimento da área de bioinformática, com diversas técnicas de sequenciamento de DNA e algoritmos para detectar inconsistências que façam uso do poder computacional do século XXI. Uma de suas principais aplicações é no estudo de desordens como câncer, síndrome de Down, doenças cardiovasculares, mal de Parkinson e mal de Alzheimer [1][2].

Há diversos tipos de variações, como SNPs, inserções, deleções, inversões, com algumas mais estudadas e fáceis de detectar do que outras. Há variações de baixa escala, que afetam apenas algumas poucas bases, e variações estruturais, que costumam afetar mais de 1000 bases.

As técnicas de detecção de variações genéticas estão sujeitas à disponibilidade de tecnologia para obtenção de dados sobre o material genético a ser examinado. Os avanços tecnológicos e algorítmicos deixam à disposição um maior volume de dados a um custo reduzido [3]. É preciso acompanhar este avanço com o desenvolvimento de técnicas que façam melhor uso destes dados, assim como estudar mais a fundo as tendências que são relevantes para a construção de modelos estatísticos úteis [3].

Além de avanços tecnológicos, a disponibilidade de dados também está sujeita a condições financeiras que determinam o sequenciamento a ser utilizado. Certas técnicas se restringem a sequenciar apenas as regiões codificantes (*exome sequencing*) que, no caso dos humanos, consiste de 1% do DNA [1]. Assim, também é interessante que haja técnicas capazes de detectar variações com um conjunto de dados restrito.

O tipo de variação explorado nesse trabalho é uma variação estrutural que tem sido alvo de muitos estudos recentes, devido à necessidade de se desenvolver técnicas que façam melhor uso do poder computacional que facilita e barateia o sequenciamento de grandes regiões de material genético; trata-se de Copy-Number Variations, ou CNVs.

Serão exploradas a fundo algumas técnicas de detecção de CNVs, que possuem vantagens e desvantagens em certos conjuntos de dados [4]. Os resultados de avaliações e comparações realizadas pelos próprios autores e outros pesquisadores serão de grande importância para se obter uma perspectiva sobre as dificuldades da área e que abordagens são mais efetivas para o problema sendo atacado [5].

Objetivo

Na área de bioinformática é bastante comum para um mesmo problema haver uma grande variedade de tipos de entrada, técnicas de obtenção de dados e diversas abordagens para se resolver o problema que se adequem a essas condições. É comum a necessidade de se escolher algoritmos adequados para casos específicos, devido em parte ao custo e tempo necessários para se obter dados e realizar experimentos.

O objetivo deste trabalho é analisar e comparar abordagens e seus resultados experimentais, a fim de alcançar um maior entendimento sobre o problema e dar uma visão geral sobre a variedade de casos reais e que métodos são desenvolvidos para trata-los. Para isso serão explorados a fundo diversos artigos que realizam tais experimentos.

Cronograma

Atividade	**Abril**				**Maio**				**Junho**				**Julho**				**Agosto**	
Levantamento bibliográfico	X	X	X	X														
Finalização da proposta de trabalho							X	X	X									
Estudo dos artigos					X	X	X	X	X	X	X	X	X	X	X	X		
Conclusão da monografia										X	X	X	X	X	X	X	X	
Elaboração da Apresentação															X	X	X	
Apresentação do trabalho																	X	X

Referências

[1] EXCAVATOR: detecting copy number variants from whole-exome sequencing data, Magi A et al, disponível em http://genomebiology.com/2013/14/10/R120
[2] Outlier-Based Identification of Copy Number Variations Using Targeted Resequencing in a Small Cohort of Patients with Tetralogy of Fallot, Bansal V et al, disponível em http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0085375
[3] Copy number variation detection using next generation sequencing read counts, Wang H et al, disponível em http://www.biomedcentral.com/1471-2105/15/109
[4] An Evaluation of Copy Number Variation Detection Tools from Whole-Exome Sequencing Data, Tan R et al, disponível em http://www.ncbi.nlm.nih.gov/pubmed/24599517
[5] Comparative Studies of Copy Number Variation Detection Methods for Next-Generation Sequencing Technologies, Duan J et al, disponível em http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0059128

Possíveis avaliadores

- Katia Silva Guimarães

- Paulo Gustavo Soares Fonseca

ASSINATURAS:

_____

Lucas Menezes Veras

Aluno

_____

Katia Silva Guimarães

Orientadora