



UNIVERSIDADE FEDERAL DE PERNAMBUCO

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CENTRO DE INFORMÁTICA

2013.2

ANÁLISE DE DADOS ALÉLICOS REAIS USANDO O PLINK E O
DIPLOTYPER

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Israel Batista Freitas da Silva (ibfs@cin.ufpe.br)

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

Recife, 29 de novembro de 2013.

Sumário

Contexto	3
Objetivos.....	4
Cronograma	5
Possíveis Avaliadores	6
Referências	7
Assinaturas	8

Contexto

Doenças renais crônicas (*Chronic Kidney Disease* – CKD¹) afetam cerca de 26 milhões de adultos nos Estados Unidos da América (EUA) (1). Além disto, CKD gera gastos substanciais nos EUA, com similar tendência global (2). Apesar do seu prevalente crescimento, o entendimento dos seus fatores de risco e o do conjunto de alterações fisiológicas associados permanecem incompletos.

Utilizando *genome-wide association studies* (GWAS), Anna e colegas identificaram suscetíveis variantes para funções renais e CKD (3). O GWAS são análises comparativas de todo o genoma de um grupo de indivíduos com característica comum, frente a uma população que não possui esta característica. Normalmente o grupo de estudo será formado por indivíduos que sofrem de uma doença comum. Este método é conhecido como caso-controle, no qual a análise da população que contém certo traço (caso) é comparada em relação aos genomas da população que não possui este traço (controle). As comparações se realizam sobre os SNPs, Polimorfismos de Nucleotídeo Simples, que são considerados a maior fonte de variabilidade genética em uma população.

Para uma investigação mais abrangente, serão realizadas análises dos SNPs usando os dados alélicos do *CKDGen Consortium*.

¹ CKD é definido como dano persistente no rim manifestado/accentuado ou pela presença de *albuminuria* ou redução de *estimated glomerular filtration* (eGFR).

Objetivos

Este trabalho de graduação tem como objetivo o estudo e análise de dados alélicos reais do CKDGen Consortium² (4).

Através de novos desafios computacionais, sobretudo por trabalhar com comparações baseadas em análises estatísticas em uma grande quantidade de dados utilizaremos o conjunto de ferramentas PLINK (5, 6) e a ferramenta Dyplotyper (7, 8). O PLINK é um software livre implementado nas linguagens de programação C e C++ com foco na análise dos dados de genótipos e fenótipos, onde os dados são processados de forma eficiente mesmo quando em grande quantidade. E o Dyplotyper que combina um novo algoritmo que clusteriza haplótipos de interesse de um conjunto de haplótipos com duas ferramentas existentes: Haploview (9), para análise de *linkage disequilibrium blocks* e haplótipos, e o PLINK para gerar todos os possíveis diplótipos de um conjunto de genótipos de amostra e calcular regressão linear e logística.

² CKDGen Consortium representa um grupo de pesquisadores dos Estados Unidos da América e da Europa que investigam a função de genes na etiologia de formas comuns em doenças renais.

Cronograma

Atividade	Novembro				Dezembro				Janeiro				Fevereiro			
Levantamento bibliográfico	X	X	X	X	X											
Conhecer PLINK usando dados simulados				X	X	X										
Utilizar PLINK em análises e testes dos dados reais					X	X	X	X	X							
Possível alteração no código do PLINK de acordo com necessidade das análises							X	X	X							
Conhecer Dyplotyper usando dados simulados								X	X	X	X					
Usar Dyplotyper em análises e testes dos dados reais									X	X	X					
Possível alteração no código do Dyplotyper de acordo com necessidade das análises										X	X					
Elaboração do relatório final										X	X	X	X	X	X	
Elaboração da apresentação														X	X	X

Possíveis Avaliadores

- Profa. Renata Maria Cardoso Rodrigues de Sousa
- Prof. Paulo Gustavo Soares da Fonseca

Referências

1. Coresh, J., Selvin, E., Stevens, L.A., Manzi, J., Kusek, J.W., Eggers, P., Van, L.F. and Levey, A.S. (2007) Prevalence of chronic kidney disease in the United States. *JAMA*, 298, 2038–2047.
2. Levey AS, Atkins R, Coresh J, Cohen EP, Collins AJ, Eckardt KU, Nahas ME, Jaber BL, Jadoul M, Levin A, Powe NR, Rossert J, Wheeler DC, Lameire N, Eknoyan G. Chronic kidney disease as a global public health problem: approaches and initiatives - a position statement from Kidney Disease Improving Global Outcomes. *Kidney Int* 2007;72:247–259. [PubMed: 17568785]
3. Köttgen, Anna, et al. "Multiple loci associated with indices of renal function and chronic kidney disease." *Nature genetics* 41.6 (2009): 712-717.
4. Fox, Caroline. CKDGen Meta-Analysis Data. Disponível em <<http://www.nhlbi.nih.gov/research/intramural/researchers/pi/fox-caroline/ckdgen-meta-analysis-data/ckdgen-meta-analysis-data.html>>. Acessado em: 28 de novembro de 2013.
5. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.
6. Purcell, Shaun. PLINK. Disponível em <<http://pngu.mgh.harvard.edu/purcell/plink/>>. Acessado em: 28 de novembro de 2013.
7. Kim, Sunshin, et al. "Diplotyper: diplotype-based association analysis." *BMC Medical Genomics* 6.2 (2013): 1-8.
8. Kim, Sunshin. Diplotype-based Association Analysis. Disponível em <<https://code.google.com/p/diplotyper/>>. Acessando em: 28 de novembro de 2013.
9. Broad Institute. Haploview. Disponível em <<http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview>>. Acessado em: 28 de novembro de 2013.

Assinaturas

Israel Batista Freitas da Silva

Orientando

Katia Silva Guimarães

Orientadora