BIOINFORMATICS

Vol. 00 no. 00 2005
Pages 1–2

genewiseR

Doc Stan

Nehil Jain [1,*], Kacper Zukowski [2] and Co-Author [2*]

[1]Department of Computer Science, Address XXXX etc.
[2]Department of XXXXXXXX, Address XXXX etc.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation:

R is the language of choice for statistical analysis in genomic research. This is because of its open source nature and the advance collection of R packages for statistical tests and visualisation of data. In GWAS studies, pooled sequencing is gaining popularity. Many tools exist to work with next-gen sequencing data in a gwas study; Bioconductor and Python are the go to places. Very few packages and tools exist for pooled seq gwas in R. This is mainly because of slow speed and memory constraints for R when working with large datasets. New high performance computing techniques in R have enabled the fullfilment of this gap. Works well on desktop and cluster configurations. Open source. Fast and intuitive methods for biology scientist to perform gwas analysis.

Results: The availablity of genewiseR on github makes it easy for scientists to install the package in their development environment is perform gwas analysis with ease. Because of addition on boostrapping feature for p-values of snps, this package allow scientist to indentify noise in their data and reduce the dataset early on in their gewas analysis.

Availability: genewiseR is open source and easily downloadable at github here. The

Contact: jain.nehil@gmail.com

1 INTRODUCTION

Genome

- what is gwas?
- why pool seq and why new skills and large datasets?
- why snps to genes?
- study design to make use of genewiseR

Text. Figure 1 shows that the above method Text Text Text Text Text Text Text. (Bag *et al*., 2001) wants to know about text follows.

$$\sum x + y = Z \tag{1}$$

*to whom correspondence should be addressed

Table 1. This is table caption

head1	head2	head3	head4
row1	row1	row1	row1
row2	row2	row2	row2
row3	row3	row3	row3
row4	row4	row4	row4

This is a footnote

Fig. 1. Caption, caption.

Fig. 2. Caption, caption.

2 DESCRIPTION

2.1 Input Data

The primary input data for this package is filtered sync files of Pooled Sequencing Data. The reads of pool-seq experiment are mapped to reference genome and then converted to multi-pileup(mpileup) format using 'Samtools' (**?**). The mpileup format is converted to sync using Popoolation2 (**?**). For detailed description refer [online documentation link here].

- Hypothesis testing, QC with bootstrapping
- Multiple testing correction
- Snps to Genes
- Calc gene level statistics
- LD assesment
- Utilities required by biologist
- Pathway analysis with Revigo/David

2.1.1 Parallelisation and Performance Sequencing pool of individuals are cheaper than individuals but its is computationally more expensive as the amount of data is multifolds. This creates some challenges associated with the desired analysis. The package is built with parallelisation in mind. Using foreach (**?**) and plyr (**?**) the time consuming analysis can take advantage of multiprocessing and cluster computing constructs.

Bofelli *et al*., 2000

© Oxford University Press 2005.

3 EXAMPLE

4 VALIDATION OR CONCLUSION

(Table 1) Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text. Bofelli *et al*., 2000 might want to know about text text text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text. Bofelli *et al*., 2000 might want to know about text text text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text.

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text. Bofelli *et al*., 2000 might want to know about text text text text

ACKNOWLEDGEMENT

Text Text Text Text Text Text Text Text Text. Bofelli *et al*., 2000 might want to know about text text text text

Funding: Text Text Text Text Text Text Text Text.

REFERENCES

Bofelli,F., Name2, Name3 (2003) Article title, *Journal Name*, **199**, 133-154.

Bag,M., Name2, Name3 (2001) Article title, *Journal Name*, **99**, 33-54.

Yoo,M.S. *et al*. (2003) Oxidative stress regulated genes in nigral dopaminergic neurnol cell: correlation with the known pathology in Parkinson's disease. *Brain Res. Mol. Brain Res.*, **110**(Suppl. 1), 76–84.

Lehmann,E.L. (1986) Chapter title. *Book Title*. Vol. 1, 2nd edn. Springer-Verlag, New York.

Crenshaw, B.,III, and Jones, W.B.,Jr (2003) The future of clinical cancer management: one tumor, one chip. *Bioinformatics*, doi:10.1093/bioinformatics/btn000.

Auhtor,A.B. *et al*. (2000) Chapter title. In Smith, A.C. (ed.), *Book Title*, 2nd edn. Publisher, Location, Vol. 1, pp. ???–???.

Bardet, G. (1920) Sur un syndrome d'obesite infantile avec polydactylie et retinite pigmentaire (contribution a l'etude des formes cliniques de l'obesite hypophysaire). PhD Thesis, name of institution, Paris, France.