

POSTER PRESENTATION

Open Access

# An integrated approach for SNP calling based on population of genomes

Nam S Vo<sup>1\*</sup>, Quang Tran<sup>2</sup>, Vinhthuy Phan<sup>1</sup>

From UT-KBRIN Bioinformatics Summit 2014  
Cadiz, KY, USA. 11-13 April 2014

## Background

The identification of genetic variants such as single nucleotide polymorphisms (SNPs) is a critical step in many applications based on NGS technologies [1]. Although many SNP calling programs have been developed, it is still challenging to accurately call SNPs, especially when coverage level is low [2]. Moreover, the determination of SNPs, which is performed through many separate steps, requires a careful selection of a diverse set of tools [3,4]. This can lead to several disadvantages, for example, one cannot incorporate information from the read alignment step into the SNP calling step or vice versa to help improve accuracy of called SNPs.

## Materials and methods

We propose a novel integrated approach to detect more true SNPs while calling fewer false positives. Different from current methods that perform read alignment and SNP calling steps separately, our method combines them methodologically to improve the accuracy of SNP identification. To effectively exploit information from a population of genomes, databases of confirmed SNPs, such as dbSNP, are employed in both aligning reads to references as well as calling SNPs. This strategy allows us to develop a novel algorithm to align reads to references that can differentiate sequencing errors from SNPs.

## Results

Based on this result, the method can call SNPs accurately and effectively even with low-coverage sequencing data. Our results on simulated data show that the method is able to call SNPs with very high precision and recall rate with low-coverage datasets.

## Conclusions

With the existence of databases of confirmed SNPs for large amounts of sequenced species, our approach provides a promising method to call accurate SNP information even with low-coverage sequencing data. This approach can also help researchers facilitate the determination of SNPs by using an integrated SNP calling tool.

## Acknowledgements

This work is partly supported by NSF CCF-1320297.

## Authors' details

<sup>1</sup>Department of Computer Science, University of Memphis, Memphis, TN 38152, USA. <sup>2</sup>Bioinformatics Program, University of Memphis, Memphis, TN 38152, USA.

Published: 29 September 2014

## References

1. Nielsen R, Paul JS, Albrechtsen A, Song YS: **Genotype and snp calling from next-generation sequencing data.** *Nat Rev Genet* 2011, **12**(6):443-451.
2. Yu X, Sun S: **Comparing a few snp calling algorithms using low-coverage sequencing data.** *BMC Bioinformatics* 2013, **14**:274.
3. Altmann A, Weber P, Bader D, Preuss M, Binder EB, Müller-Myhsok B: **A beginners guide to snp calling from high-throughput dna-sequencing data.** *Hum Genet* 2012, **131**(10):1541-1554.
4. Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efreanova M, Krabichler B, Speicher MR, Zschocke J, Trajanoski Z: **A survey of tools for variant analysis of next-generation genome sequencing data.** *Brief Bioinform* 2014, **15**(2):256-278.

doi:10.1186/1471-2105-15-S10-P30

**Cite this article as:** Vo et al.: An integrated approach for SNP calling based on population of genomes. *BMC Bioinformatics* 2014 **15**(Suppl 10):P30.

\* Correspondence: nsvo1@memphis.edu

<sup>1</sup>Department of Computer Science, University of Memphis, Memphis, TN 38152, USA

Full list of author information is available at the end of the article