World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:10, No:09, 2016

Efficient Reuse of Exome Sequencing Data for Copy Number Variation Callings

Authors : Chen Wang, Jared Evans, Yan Asmann

Abstract : With the quick evolvement of next-generation sequencing techniques, whole-exome or exome-panel data have become a cost-effective way for detection of small exonic mutations, but there has been a growing desire to accurately detect copy number variations (CNVs) as well. In order to address this research and clinical needs, we developed a sequencing coverage pattern-based method not only for copy number detections, data integrity checks, CNV calling, and visualization reports. The developed methodologies include complete automation to increase usability, genome content-coverage bias correction, CNV segmentation, data quality reports, and publication quality images. Automatic identification and removal of poor quality outlier samples were made automatically. Multiple experimental batches were routinely detected and further reduced for a clean subset of samples before analysis. Algorithm improvements were also made to improve somatic CNV detection as well as germline CNV detection in trio family. Additionally, a set of utilities was included to facilitate users for producing CNV plots in focused genes of interest. We demonstrate the somatic CNV enhancements by accurately detecting CNVs in whole exome-wide data from the cancer genome atlas cancer samples and a lymphoma case study with paired tumor and normal samples. We also showed our efficient reuses of existing exome sequencing data, for improved germline CNV calling in a family of the trio from the phase-III study of 1000 Genome to detect CNVs with various modes of inheritance. The performance of the developed method is evaluated by comparing CNV calling results with results from other orthogonal copy number platforms. Through our case studies, reuses of exome sequencing data for calling CNVs have several noticeable functionalities, including a better quality control for exome sequencing data, improved joint analysis with single nucleotide variant calls, and novel genomic discovery of under-utilized existing whole exome and custom exome panel data.

Keywords : bioinformatics, computational genetics, copy number variations, data reuse, exome sequencing, next generation sequencing
Conference Title : ICBCBBE 2016 : International Conference on Bioinformatics, Computational Biology and Biomedical Engineering
Conference Location : Chicago, United States
Conference Dates : September 19-20, 2016