Inferring Copy Number Variation Networks From The Qatari Genome

10.5339/qfarc.2014.HBPP1070

Noha A. Yousri, Ph.d.; Khalid A. Fakhro; Ronald G. Crystal; Karsten Subre

Corresponding author:
nay2005@qatar-med.cornell.edu
Weill Cornell Medical College Qatar, Doha, Qatar

Abstract

Background: Thousands of Copy Number Variations (CNVs) obtained from Next Generation Sequencing (NGS) technologies present a rich information for biologists. Such huge information is tempting for discovering the inherent characteristics for a population’s genome. Copy Number Variations (CNVs) are deletions and duplications in the genome, that may associate with certain phenotypic characteristics as diseases or may be specific to populations.

Objectives: Developing a method for identifying relations between CNVs in the Qatari population. This is used for detecting networks of CNVs that span several pathways, and which are enriched in deletions or duplications for this population. It is also used to investigate genes and pathways that are affected by multiple CNVRs in the same individual/population. Finding networks associated with each Qatari sub-population (Bedouin, Persian and African) are also investigated.

Materials & Methods: A set of 108 Qatari genomes was used in this study. A method for inferring Copy Number Variations Networks (CNVN)s using genes associated with CNVs is proposed. After reducing CNVs to Copy Number Variation Regions (CNVRs), the CNVRs are annotated with associated genes using biological databases. Biological pathways associated with those genes are then identified. For each individual, pairwise similarities between CNVRs are calculated based on the number of overlapping gene pathways. Similarity values are used to construct the edges of an individual-level CNVN. Aggregation of edges from all individuals' CNVNs is used to construct a population-level CNVN. Based on a specific edge weight threshold, sub-networks that connect groups of related CNVRs are found.

Results: A set of 108 genomes was used to investigate CNVNs associated with the 3 subpopulations (Bedouin, Persian and African) in the Qatari cohort. More than 16,000 CNVRs were used to construct the network, based on around 3000 genes associated with the CNVRs, and around 200 pathways associated with those genes. Sub-networks connecting CNVRs from several chromosomes were identified.