Predicting Susceptibility to Coronary Artery Disease using Single Nucleotide Polymorphisms with a Large-Scale Data Extraction from PubMed and Validation in an Asian Population Subset

Authors : K. H. Reeta, Bhavana Prasher, Mitali Mukerji, Dhwani Dholakia, Sangeeta Khanna, Archana Vats, Shivam Pandey, Sandeep Seth, Subir Kumar Maulik

Abstract : Introduction Research has demonstrated a connection between coronary artery disease (CAD) and genetics. We did a deep literature mining using both bioinformatics and manual efforts to identify the susceptible polymorphisms in coronary artery disease. Further, the study sought to validate these findings in an Asian population. Methodology In first phase, we used an automated pipeline which organizes and presents structured information on SNPs, Population and Diseases. The information was obtained by applying Natural Language Processing (NLP) techniques to approximately 28 million PubMed abstracts. To accomplish this, we utilized Python scripts to extract and curate disease-related data, filter out false positives, and categorize them into 24 hierarchical groups using named Entity Recognition (NER) algorithms. From the extensive research conducted, a total of 466 unique PubMed Identifiers (PMIDs) and 694 Single Nucleotide Polymorphisms (SNPs) related to coronary artery disease (CAD) were identified. To refine the selection process, a thorough manual examination of all the studies was carried out. Specifically, SNPs that demonstrated susceptibility to CAD and exhibited a positive Odds Ratio (OR) were selected, and a final pool of 324 SNPs was compiled. The next phase involved validating the identified SNPs in DNA samples of 96 CAD patients and 37 healthy controls from Indian population using Global Screening Array. ResultsThe results exhibited out of 324, only 108 SNPs were expressed, further 4 SNPs showed significant difference of minor allele frequency in cases and controls. These were rs187238 of IL-18 gene, rs731236 of VDR gene, rs11556218 of IL16 gene and rs5882 of CETP gene. Prior researches have reported association of these SNPs with various pathways like endothelial damage, susceptibility of vitamin D receptor (VDR) polymorphisms, and reduction of HDL-cholesterol levels, ultimately leading to the development of CAD. Among these, only rs731236 had been studied in Indian population and that too in diabetes and vitamin D deficiency. For the first time, these SNPs were reported to be associated with CAD in Indian population. Conclusion: This pool of 324 SNP s is a unique kind of resource that can help to uncover risk associations in CAD. Here, we validated in Indian population. Further, validation in different populations may offer valuable insights and contribute to the development of a screening tool and may help in enabling the implementation of primary prevention strategies targeted at the vulnerable population.

Keywords : coronary artery disease, single nucleotide polymorphism, susceptible SNP, bioinformatics
Conference Title : ICCCS 2023 : International Conference on Cardiology and Cardiac Surgery
Conference Location : Singapore, Singapore
Conference Dates : September 04-05, 2023