MATCHVAR: An Efficient Online Tool for Gene Variants Annotation from High throughput Sequencing Data

Bingbo Zhou
State Key Laboratory of Digital Medical Engineering, School of Biological Science and Medical Engineering, Southeast University, Nanjing, PR China
The Center of Medicine Genetic in Gansu Provincial Maternity and Child-care Hospital, Gansu Provincial Clinical Research Center for Birth Defects and Rare Diseases, Lanzhou, PR China

Chuan Zhang
The Center of Medicine Genetic in Gansu Provincial Maternity and Child-care Hospital, Gansu Provincial Clinical Research Center for Birth Defects and Rare Diseases, Lanzhou, PR China

Xue Chen
The Center of Medicine Genetic in Gansu Provincial Maternity and Child-care Hospital, Gansu Provincial Clinical Research Center for Birth Defects and Rare Diseases, Lanzhou, PR China

Jing Li
The Reproductive Medicine Center II, Gansu Provincial Maternity and Child-care Hospital, Lanzhou, PR China

Yunfei Bai *
State Key Laboratory of Digital Medical Engineering, School of Biological Science and Medical Engineering, Southeast University, Nanjing, PR China

Abstract

The maturation and extensive clinical application of Next-Generation Sequencing (NGS) technologies have yielded a substantial volume of genetic variation data throughout the human genome, and genome annotation assumes a pivotal position in identifying genetic variants and in conducting other characterization studies. However, accurately identifying meaningful genetic variants from this large dataset remains a significant challenge with existing methods. To address this, we developed Gene Variation Simulator and Annotator (GVSAA), an online tool for variation simulation and functionally annotating single nucleotide variants (SNVs) and insertion/deletions (Indels) from NGS data. MATCHVAR offers multiple key annotation functions as part of the functional annotation section of GVSAA, including annotation based on gene information from MANE-indexed transcripts, population frequency and software prediction databases. MATCHVAR also supports the upload of custom (in-house) databases, providing flexibility in data sources. For rapid annotation of a small number of variants, we also developed a Cloud-based annotation approach using the Ensemble public API, which ensures efficient use of multiple functional annotations for these variants. To validate local annotation utility, we tested the SNV and indel of human genomes with different numbers of variations using MATCHVAR. The results show that MATCHVAR exhibits significant performance in annotation tasks. MATCHVAR is freely available at https://matchvar.intelligene.cn.

Keywords

MATCHVAR, GVSAA, Next-Generation Sequencing, Genetic Variants, Data simulation, VEP Cloud.