第２６５回原医研セミナー
第２９回放射線災害・医科学研究 機構・拠点研究推進ミーティング

Zoom( )ID 890 6191 5257

Zoom URL

https://us02web.zoom.us/j/89061915257?pwd=Uk93L2JWWDJ3dnFkYmkvSjFGN21DZz09

Zoom 538773 URL

Machine Learning-based Quality Control Procedure For Merging Whole Genome Sequencing Datasets In Rare Genetic Variant Research.

Uladzislau KORZUN

Increased availability of Whole-genome sequencing (WGS) resulted in creation of numerous sequencing projects and genomic information databanks. By combining datasets provided by different institutions one can easily increase the sample size of a study while avoiding direct costs of sequencing these samples. Unfortunately, sequencing data obtained from different institution possesses a unique pattern of technical errors, represented as genetic variants of artifactual nature. These variants of nonbiological origin compromise the results of a subsequent association study. Quality control (QC) of WGS data is essential to prevent false associations. Traditional approaches mitigate artifactual variants by removing rare genetic variants from combined datasets. This practice hinders the investigation of rare and less-common genetic variants in phenotypic trait inheritance. We are working on a machine learning-based QC procedure that reduces artifactual variants while preserving true rare genetic variation.

2023 9

2024 4 5 web

0–100 0= 100=

082-257-5802 5802