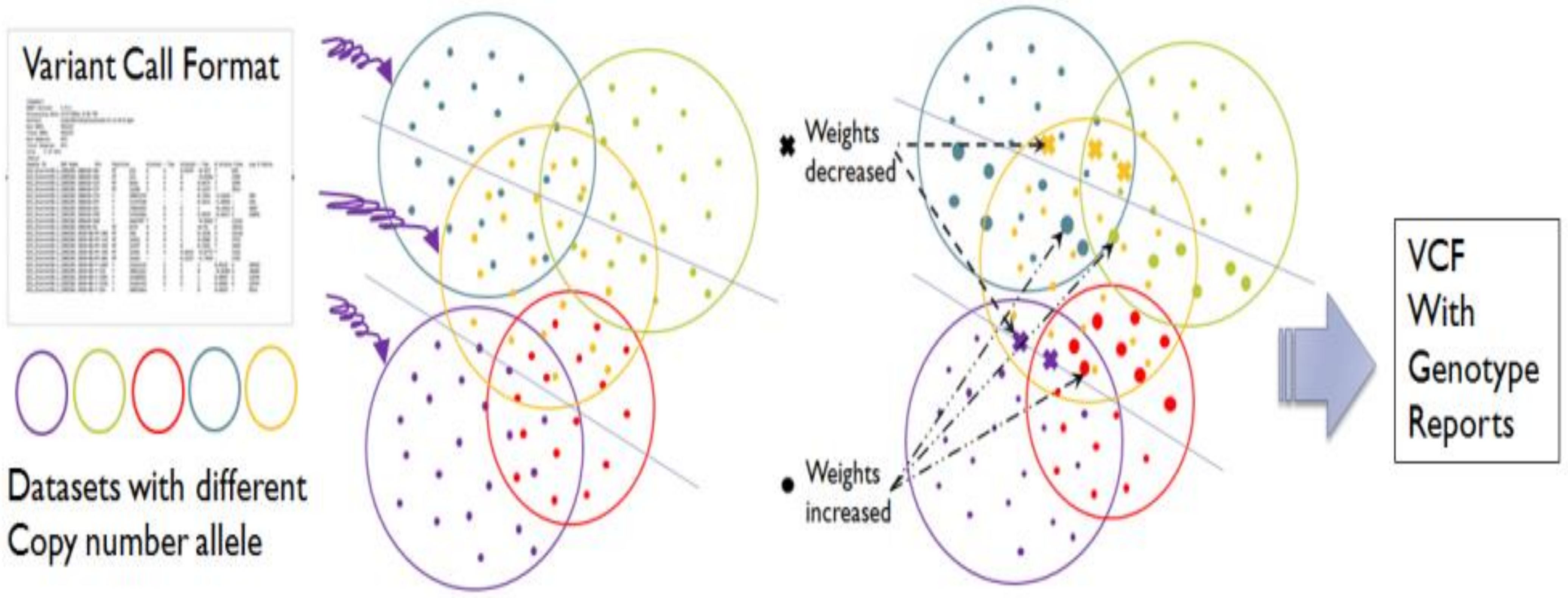
A structural variation genotyping algorithm enhanced by CNV quantitative transfer

Tian ZHENG, Xinyang QIAN, Jiayin WANG

Frontiers of Computer Science, DOI: 10.1007/s11704-021-1177-z



- We propose a transfer learning-based method to genotype structural variations accurately considering CNVs. The method first divides the instances with different allelic copy numbers and trains the basic machine learning framework with different genotype datasets. By adjusting the weights of the instances with different allelic copy numbers, the contribution of all the instances to genotyping can be maximized, and the genotyping errors of heterozygote variations caused by CNVs can be minimized.

- Different CNV examples have positive and negative effects on the correct typing, and the degree is different. The larger the copy number is, the easier it is to lead to misjudgment.
- This paper proposes a case-based transfer learning model, which gives different weights to the cases with different copy numbers, and dynamically adjusts them in the training process, which can promote the case weight of correct typing to increase and the case weight of interference typing to decrease, so as to realize the genotyping algorithm without the interference of CNVs.
- We carried out experiments on the simulated data and the real data of KGP. The overall accuracy of the algorithm is 10-20% higher than those of popular methods such as ***GATK, Gindel*** and ***Facets.***