# CircMarker: A Fast and Accurate Algorithm for Circular RNA Detection

Xin Li, Chong Chu, Jingwen Pei, Ion Măndoiu, and Yufeng Wu*

Computer Science & Engineering Dept., University of Connecticut, Storrs, CT, USA
{xin.li,chong.chu,jingwen.pei,ion.mandoiu,yufeng.wu}@uconn.edu

## 1 Abstract

While RNA is often created from linear splicing during transcription, circular RNA (or circRNA) is a type of RNA which forms a covalently closed continuous loop. It is now believed that circRNA plays an important biological role in diseases and traits. Several experimental methods, such as RNase R, have been designed to enrich circRNA while degrade linear RNA. Although several useful software tools for circRNA detection have been developed as well, these tools may miss many circular RNA. Also, existing tools are slow for large data because they often depend on reads mapping. To deal with these problems, we developed a new computational approach, named CircMarker, based on k-mers rather than reads mapping for circRNA detection. CircMarker takes advantage of transcriptome annotation files to create space-efficient k-mer table, and applies several different criteria and filters for circRNA detection. We also compared CircMarker with other three circRNA detection tools in both simulation and real data. Two different circRNA coverage cases have been applied for simulation data and CircMarker performs the best in all indicates of these two cases, including the number of called circRNA, accuracy, and running time. For real data evaluation, two different strategies have been applied. The first strategy uses RNase R treated sequence reads with public database, and the second one uses both RNase R Treated/Untreated Data from one raw specimen. Empirical results show that CircMarker can find more reliable circular RNA and obtain higher reliability ratio and consensus-based sensitivity with low bias. CircMarker is generally 5 times faster than others.

**Keywords:** circular RNA, high-throughput sequencing, genomics, RNA-Seq