XIN LI

Storrs, Mansfield, CT, 06269 ⋄ Website: lxwgcool.github.io
Phone: 860-617-9785 ⋄ Email: xin.li@uconn.edu

EDUCATION

University of Connecticut, Storrs   *2014 - 2019*
Ph.D. in Computer Science
Beihang University, Beijing   *2006 - 2009*
M.S. in Computer Software Engineering

TECHNICAL SKILLS

- Proficient in C++ (10 years); skillful in Linux shell, Python, C#

- Proficient in high-performance computing (HPC), Slurm

- Proficient in DNA/RNA sequence (NGS) analysis and STR fragment analysis

- Skillful in software design, machine learning, bioinformatics and statistics

RESEARCH PROJECTS

Detecting genomic deletions from NGS data with unsupervised learning
Develop a new method called EigenDel. EigenDel first uses discordant & clipped reads to get initial deletion candidates, and then it clusters similar candidates by using PCA and hierarchical clustering. Finally, EigenDel uses a carefully designed approach for calling true deletions from each cluster.

Detecting circular RNA from high-throughput sequence data with de Bruijn graph
Develop a new method named CircDBG. CircDBG creates a de Bruijn graph based on k-mers from exons, and then it takes advantage of this graph to find the relationship between k-mer of reads and donor/acceptor exon for circRNA detection. A special case of circRNA, chimeric circRNA, is reported.

CircMarker: a fast and accurate algorithm for circular RNA detection
Develop a new computational approach named CircMarker. CircMarker is based on k-mers rather than reads mapping and takes advantage of annotation files to create the k-mer table for circRNA detection.

WORK EXPERIENCE

HPC Administrator   2016 - present
University of Connecticut, Storrs, CT

- Provide assistance to users of HPC cluster in resolving issues and managing support tickets

- System performance analysis and tuning; monitor, analyze, and correct system issues

- Build, install and support user requested software; create singularity image for kernel sensitive tools, such as tensorflow (CUDA-based version)

- Deploy new hardware, such as computing node, GPU node and clustered file system

Research Assistant   2014 - present
University of Connecticut, Storrs, CT

- Design computational algorithm for complex genome analysis based on NGS sequencing data

- DNA/RNA sequence analysis in multi-research fields, such as SV deletion, circRNA and gap filling

- Build pipeline to process NGS data by using 3rd-party tools, such as BWA, SAMTools and Picard

R&D Director 2011 - 2014
Software Engineer 2009 - 2014
Todaysoft Ltd, Beijing, China (collaborator of SoftGenetics LLC, State College, PA)

- Lead R&D of DNA fragment analysis software (C++)

- Software architecture design, algorithm research, and core code implementation

- Requirements analysis and prototype development for the market in both China and US

- Major Products: ChimerMarker (chimerism analysis), GeneMarker HID (STR Human Identity), GeneMarker (genotype analysis), GeneMarker MTP (multi-template processor), GeneMarker Melting Curve (thalassemia and tuberculosis resistance), GeneMarker Maize (corn breeding)

Research Intern 2007 - 2008
Siemens Corporate Technology, Beijing, China

- Software design, algorithm research and implementation in computer graphics

- Develop a product: 3D simulation system of steel refrigeration conveyor line (C#)

AWARDS AND PATENT

Doctoral Dissertation Fellowship	2019
Predoctoral Fellowship	2019
Predoctoral Honorable Mention	2017, 2018
Travel fellowship, ICCABS	2017
Travel fellowship, ISBRA	2017
Jin Hui Prize, Science and Technology Department of Jiangsu Province, China	2013
PATENT: Wavelet Method to Process the Melting Curve	2013

ACADEMIC ACTIVITIES

Poster in 17th APBC	2019
Paper presentation in 13th and 14th ISBRA	2017, 2018
Poster in RECOMB	2017, 2018
Paper presentation in 7th ICCABS	2017
Exhibitor in ASHG	2010

PUBLICATIONS

1. Detecting genomic deletions from high-throughput sequence data with unsupervised learning. Xin Li, Yufeng Wu, *Scientific Report*, Revise and Resubmit, 2019

2. Detecting circular RNA from high-throughput sequence data with de Bruijn graph. Xin Li, Yufeng Wu, *BMC Genomics*, Accepted, 2019

3. CircMarker: a fast and accurate algorithm for circular RNA detection. Xin Li, Chong Chu, Jingwen Pei, Ion Mandoiu, Yufeng Wu, *BMC Genomics*, 2018

4. CLADES: A classification-based machine learning method for species delimitation from population genetic data. JW Pei, Chong Chu, Xin Li, Bin Lu, YF Wu, *Molecular Ecology Resources*, 2018

5. GAPPadder: a sensitive approach for closing gaps on draft genomes with short sequence reads. Chong Chu, Xin Li and Yufeng Wu, *BMC Genomics*, 2017

6. SpliceJumper: a classification-based approach for calling splicing junctions from RNA-seq data. Chong Chu, Xin Li and Yufeng Wu, *BMC Bioinformatics*, 2017