

## **Jianying Li**

3405 Rockland Ridge Drive

Cary, NC 27519

Work: (919) 843-9191

Cell: (919) 607-1968

Email: [jianying.li@gmail.com](mailto:jianying.li@gmail.com)

### **OBJECTIVE**

Seek a statistics/bioinformatics position in an academic research area

### **QUALIFICATION**

- Professionally trained in bioinformatics with **nine-year** working/research experiences
- **Two-year** hands-on experiences in high-throughput genome informatics analysis with focus on Next Generation Sequencing (NGS)
- Strong in statistical modeling and statistical/computational programming in **SAS, JMP, R, Perl, Java**, and **C++**
- A highly productive researcher with peer-reviewed papers on highly regarded journal
- Self-motivated individual with smooth communication skills

### **EDUCATION**

Master of Bioinformatics, North Carolina State University, 2002

Master of Science in Food Microbiology, North Carolina State University, 2001

B.S. in Microbiology, Nankai University Tianjin, 1990

### **WORKING EXPERIENCE**

#### **UNC Chapel Hill, Lineberger Comprehensive Cancer Center**

10/2008 to present      Senior Bioinformatician

#### ■ Selected recent projects involved

- Leading and participating in establishing in-house pipeline for RNA-seq project. I am a main developer for RNAseq QC modules and analytical workflow in SeqWare ([http://sourceforge.net/apps/mediawiki/seqware/index.php?title=Main\\_Page](http://sourceforge.net/apps/mediawiki/seqware/index.php?title=Main_Page)). This project involves benchmarking and testing open source software package, implementing in-house statistical analysis, constructing customized database for data annotation, building LIMS and data storage database for *ad hoc* analyses. All statistical components are implemented in **R**, sequence processing and manipulation is implanted in **Perl**, and the final software/application is wrapped with **Java**

- Lead author on the Principal Variance Component Analysis (**PVCA**) procedure leveraging mixed linear model for genetic/genomic data analysis. The theory was published as a book chapter and the model was developed both in **SAS** (currently an analytical procedure “AP” in JMP Genomics software package from SAS Institute) and **R** (<http://www.niehs.nih.gov/research/resources/software/pvca/index.cfm>) for open source implementation.
- Participating in recently funded (by NIH) **The Cancer Genome Atlas (TCGA)** project, mainly relying on next generation sequence platform (plus genotyping and other high throughput method) to detect new mutation in cancer cell, find new transcripts, and study novel cancer mechanistic system. My main responsibility is statistical analysis and model on high-throughput NGS data.
- Collaborating and leading the bioinformatics component on a project “to determine the DNA replication system for cancer mechanism using next generation whole genome sequencing”

**SRA International, Inc.,** Durham, NC, 07/2008 to 10/2008 Senior Bioinformatics Developer

- Provided general bioinformatics/biostatistics support for DIR scientists at NIEHS
- Provided support through collaborated research projects with DIR investigators
- Provided support through independently initiated research projects by developing new methodologies, new algorithms, and new applications to facilitate the DIR research

**Lockheed Martin,** 01/2003 to 10/2008 Bioinformatic Scientist

- Bioinformatics application and statistical methodology development
- Developed in house scientific software in **Java**, for microarray quality control of in house RNA quality monitoring, hybridization quality assessment, data acquisition software performance evaluation
- Developed in house scientific software in **C++** implementing Fuzzy Logic algorithm for Rat acetaminophen dosing project
- Developed in house software in **R** providing analytical component in an in-house developed Oracle database Chemical Effect Biological System (CEBS)
- Implemented Mixed Linear Model in **SAS** for customized Micro Array data analysis, include data quality evaluation, data normalization, and statistical modeling
- Statistical consultation and bioinformatics support
- Provided training of application statistics using **JMP**: acted as a leading role to coordinate with IT training department to design class, construct class materials and present in a live training class
- Performed regular microarray gene expression and proteomic data analysis with commercial software package, freeware as well as in house developed software in **Java**, **R**, **SAS** and **JMP**

■ Selected research projects involved

- Developed statistic prediction models in the FDA MAQC consortium for standardized procedure for prediction phenotypic outcome using the transcriptome biomarkers, and currently involving in cross-platform and cross-tissue prediction research
- Participated in scientific director's CpG project, implementing genome searching algorithms to search for domains in human and mouse genome
- Participated in the collaborated research for statistic modeling in Toxicogenomics Research Consortium cross platform and cross laboratory gene expression

**University of Hawaii/North Carolina State University, 1998-2002 Research Assistant**

- Research on genes (in *Listeria monocytogenes*) responsible for surface antigen determinants involving all molecular technique, i.e. DNA/RNA manipulation, microarray experiment, and function study
- Research involving search for major genes contributing to favorable commercial properties in loblolly pine

**Tianjing Import and Export Commodity Inspection Bureau of the People's Republic of China (PRC), 1990 – 1998**

- Implemented and executed national/international standardization systems include: GB, IB, ISO9000, and GMP/HACCP, to ensure laboratory performance excellence
- Supervised personnel in routine laboratory inspections and researches under budget and ensured excellent laboratory performance
- Performed general food microbiological testing and research

**HARDWARE AND SOFTWARE SKILLS**

- Operating systems: **Windows, Linux, and Unix**
- Programming languages: **Java, C++, SQL, Perl, SAS, R, JMP**
- Database system: MySQL, PostgreSQL
- Genetic/genomic analysis package: JMP Genomics Suite, Rosetta Resolver, GeneSpring, Bioconductor package, IPA, GoMiner
- NextGenSeq data analysis package: MAQ, BWA, Tophat/Bowtie/Cufflinks, Samtools, SeqWare, Galaxy, FASTX-toolkit etc.

## SELECTED RECENT PUBLICATIONS

J Huang, W Shi, J Zhang, J W Chou, R S Paules, K Gerrish, **J Li**, et al on the MAQC consortium, (2010) Genomic indicators in the blood predict a drug-induced liver injury response **The Pharmacogenomics Journal** (2010) 10, 267–277

**Co-author** on the MAQC consortium, (2010) “The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models” **Nature Biotechnology** (advance online access: 10.1038/nbt.1665)

Cabanski CR, Qi Y, Yin X, Bair E, Hayward MC, **Li J**, et al. (2010) SWISS MADE: Standardized WithIn class Sum of Squares to evaluate Methodologies And Dataset Elements **PloS ONE** 5(3): e9905

**Jianying Li**, Chu-TZ, Wolfinger, R, Bushel, P.R. (2009) Batch Effects and Noise in Microarray Experiments: Sources and Solutions. John Wiley & Sons, Ltd.

Song, M.O., **J. Li**, and J.H. Freedman (2009) Physiological and Toxicological Transcriptome Exposure Levels Changes in HepG2 Cells Exposed to Copper” **Physiol. Genomics**. 38: 386-401.

P. R. Bushel, A. N. Heinloth, **J. Li**, et al (2007) “Blood Gene Expression Signatures Predict Exposure Levels” **PNAS**. v.104, no. 46: 18211-18216

**Co-author** on (2005) “Standardizing global gene expression analysis between laboratories and across platforms” **Nature Methods** 2:351-356

**Co-author** on (2004) “Empirical evaluation of data transformations and ranking statistics for microarray analysis” **Nucleic Acids Res**. 32: 5471 – 5479

## REFERENCES

References are available upon request