

**BIOGRAPHICAL SKETCH**

Provide the following information for the Senior/key personnel and other significant contributors.  
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Qingguo Wang

eRA COMMONS USER NAME (credential, e.g., agency login): QINGGUO\_WANG

POSITION TITLE: Professor of Computer Science and Data Science

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Chengdu University of Technology, Chengdu, China	B.S.	07/1998	Nuclear Technology
Shanghai Jiao Tong University, Shanghai, China	M.S.	03/2004	Computer Application
University of Missouri, Columbia, MO	M.S.	05/2010	Computer Science
University of Missouri, Columbia, MO	Ph.D.	05/2011	Computer Science
Vanderbilt University, Nashville, TN	Postdoctoral	08/2014	Computational Biology
Memorial Sloan Kettering Cancer Center, New York, NY	Postdoctoral	09/2014	Computational Biology

**A. Personal Statement**

My research is in the interdisciplinary field of Machine Learning, Computational Biology, and Cancer Genomics. I graduated from the University of Missouri with a PhD in Computer Science, where I worked primarily on machine learning algorithms for predicting protein 3D structures. The algorithms we created helped our Missouri team win the top honors in the international 9th CASP (Critical Assessment of Techniques for Protein Structure Prediction) competition. Besides research, I am also committed to training students machine learning and data analyst skills, in particular after joining Lipscomb University as a faculty member in 2016. Numerous projects that we developed in classroom were published in journals or proceedings of international conferences, with students as first or co-authors.

My training in Computational Biology and Cancer Genomics was received from Vanderbilt University and Memorial Sloan Kettering Cancer Center (MSKCC). At Vanderbilt, as a postdoctoral fellow, I focused on applying next generation sequencing (NGS) data to identify cancer- or drug resistance-causing variants in tumor genomes. I developed several popular tools at the time to help people find needles in big NGS data. One is VirusFinder, the first fully automatic software for characterizing integration sites of undiagnosed viruses of arbitrary types through the NGS data. VirusFinder is widely used due to its uniqueness and accuracy. In 2014, I joined MSKCC, where with my expertise in computing and NGS I contributed to several large-scale cancer studies, e.g. The Cancer Genome Atlas Prostate Adenocarcinoma (TCGA- PRAD), for which I used RNA sequencing data to detect an important variant in human gene Androgen Receptor. My results were incorporated in a highly cited TCGA marker paper published in *Cell* journal in 2015.

1. The Cancer Genome Atlas Research Network. (2015). The molecular taxonomy of primary prostate cancer. *Cell*, 163(4), 1011-25. PMID: PMC4695400
2. Wang Q, Armenia J, Zhang C, Penson AV, Reznik E, Zhang L, Minet T, Ochoa A, Gross BE, Iacobuzio-Donahue CA, Betel D, Taylor BS, Gao J and Schultz N. (2018). Unifying cancer and normal RNA sequencing data from different sources. *Scientific Data*, 5, 180061. PMID: PMC5903355

## B. Positions and Honors

### Positions and Employment

1998-2001	Engineer	Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai, China
2001-2004	Research Assistant	Shanghai Jiao Tong University, Shanghai, China
2004-2004	Software Engineer	Shanghai Hewlett-Packard Co., Ltd, Shanghai, China
2005-2005	Senior Software Engineer	Shanghai Digital Intelligence System Tech. Co., Ltd., Shanghai, China
2006-2011	Research Assistant	Computer Science Department, University of Missouri, MO
2011-2014	Postdoctoral Fellow	Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN
2014-2014	Postdoctoral Fellow	Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY
2014-2016	Computational Engineer III	Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY
2016-2020	Associate Professor	College of Computing & Technology, Lipscomb University, Nashville, TN
2021-present	Professor	School of Applied Computational Sciences, Meharry Medical College, Nashville, TN

### Honors

1995	First place, Mathematical Contest in Modeling, Chengdu University of Technology, China.
1996	Second place, National Collegiate Mathematical Contest in Modeling, China
1997	Second place, National Collegiate Electronics Design Contest in Sichuan Province, China
1997	Outstanding Undergraduate Student Scholarship of Sichuan Province, China
1998	Outstanding Graduate Student Award of Sichuan Province, China
2008	Shumaker Fellowship, University of Missouri, MO
2009	2nd place, Computer Science Department Poster Competition, University of Missouri, MO
2010	Upsilon Pi Epsilon (UPE) Scholarship Award - UPE is the only international honor society for the computing and information disciplines, MO
2010	First place in the human prediction for protein 3D structures in the worldwide competition CASP9 and top two places in global model quality assessment in CASP9, MO
2011	Outstanding Student Award, College of Engineering, University of Missouri, MO
2011	Nominated for Donald K. Anderson GRA Award (1 per year at the University of Missouri)
2013	Our paper for VirusFinder is among the top 10% most cited PLOS ONE articles
2017	Researcher of the Year Award, College of Computing & Technology, Lipscomb University, TN
2017	The UPE chapter I supervised won Outstanding Chapter Award in annual national convention
2018	Our paper in a short list of candidates to win best poster award in HEALTHINF conference
2019	Dean's Award, College of Computing & Technology, Lipscomb University, TN
2017 – 2020	Helped students win UPE Scholarship Awards four years in a row

## C. Contributions to Science

### 1. Contributions to cancer studies

I have participated in many collaborative cancer studies, including The Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD). I helped analyze the sequencing data in these studies. For the TCGA-PRAD project, I processed the RNA sequencing data of prostate cancer samples to detect an important variant in human gene Androgen Receptor. My results were incorporated in a highly cited TCGA marker paper published in *Cell* journal in 2015.

- a. The Cancer Genome Atlas Research Network. (2015). The molecular taxonomy of primary prostate cancer. *Cell*, 163(4), 1011-25. PMID: PMC4695400

- b. Thomas LR, Wang Q, Grieb BC, Phan J, Sun Q, Olejniczak ET, Clark T, Dey S, Foshage AM, Lorey S, Howard GC, Cawthon B, Ess KC, Eischen CM, Zhao Z, Fesik SW and Tansey WP. (2015). Interaction with WDR5 promotes target gene recognition and tumorigenesis by MYC. *Molecular Cell*, 58(3), 440-452. PMID: PMC4427524
- c. Liu Y, Chen C, Xu Z, Scuoppo C, Rillaan CD, Gao J, Spitzer B, Bosbach B, Kastenhuber ER, Baslan T, Ackermann S, Cheng L, Wang Q, Niu T, Schultz N, Levine RL, Mills AA, and Lowe SW. (2016). Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. *Nature*, 531(7595), 471-475. PMID: PMC4836395
- d. Reznik E, Wang Q, La K, Schultz N and Sander C. (2017). Mitochondrial Respiratory Gene Expression is Suppressed in Many Cancers. *eLife*, 6, e21592. PMID: PMC5243113
- e. Gallant JN, Sewell A, Almodovar K, Wang Q, Dahlman KB, Abramson RG, Kapp ME, Brown BT, Boyd KL, Gilbert J, Cohen DN, Yarbrough WG, Zhao Z and Lovly CM. (2019). Genomic landscape of a rare metastatic malignant proliferating tricholemmal tumor and its response to PI3K inhibition. *Npj Precision Oncology*, 3(1), 5. PMID: PMC6377617

## **2. Contributions to Computational Biology**

We have created several software tools for analyzing next generation sequencing data from cancer patients. The open-source software VirusFinder developed by me is the first fully automatic tool for characterizing integration sites of undiagnosed viruses of arbitrary types through the sequencing data. It has been used widely by researchers in the field to detect virus integration sites in tumor genomes. Accordingly, the article for VirusFinder is among the top 10% most cited ones published by the journal PLOS ONE. Below are the descriptions and URLs of four recent tools I developed or contributed to.

- a. NGS Catalog: a database I created for collecting NGS studies and their mutation characteristics.  
URL of the database: <https://bioinfo.uth.edu/NGS/index.html?csrt=11534114009283502775>
- b. VirusFinder: open-source tool I created for detecting viruses and their insertion sites in human genomes.  
URL of the source code: <https://bioinfo.uth.edu/VirusFinder/>
- c. MAF2MAF: a program I contributed to for cleaning mutation data and for converting from/to MAF files.  
URL of the source code: <https://github.com/mskcc/vcf2maf/blob/main/maf2maf.pl>
- d. RNAseqDB: open-source software I created for processing, correcting, and integrating RNA-seq data.  
URL of the source code: <https://github.com/mskcc/RNAseqDB>

The corresponding publications are provided below,

- a. Xia J, Wang Q, Jia P, Wang B, Pao W and Zhao Z. (2012). NGS catalog: A database of next generation sequencing studies in humans. *Human mutation*, 33(6), E2341-55. PMID: PMC4431973
- b. Wang Q, Jia P and Zhao Z. (2013). VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. *PLoS ONE*, 8(5), e64465. PMID: PMC3663743
- c. Wang Q, Jia P and Zhao Z. (2015). VERSE: a novel approach to detect virus integration in host genomes through reference genome customization. *Genome Medicine*, 7(1), 1-9. PMID: PMC4333248
- d. Wang Q, Armenia J, Zhang C, Penson AV, Reznik E, Zhang L, Minet T, Ochoa A, Gross BE, Iacobuzio-Donahue CA, Betel D, Taylor BS, Gao J and Schultz N. (2018). Unifying cancer and normal RNA sequencing data from different sources. *Scientific Data*, 5, 180061. PMID: PMC5903355

## **3. Development of machine learning algorithms**

We have also created many innovative algorithms using machine learning to analyze various types of data. Below are several examples (the last four are from my former students).

- a. Wang Q, Shang C, Xu D, and Shang Y. (2013). New MDS and clustering based algorithms for protein model quality assessment and selection. *International Journal on Artificial Intelligence Tools*, 22(5), 1360006
- b. Pigman M, Le H, Bhagat U, Thompson M and Wang Q. (2018). Analyzing Inventory Data Using K-Means Clustering. *Proceedings of the International Conference on Data Science (ICDATA'18)*, 117-122

