

Wei Zhang

CONTACT INFORMATION	Computer Science and Engineering 200 Union Street SE 6-240 University of Minnesota Twin Cities, MN 55455	Phone: (507)313-0967 http://www-users.cs.umn.edu/~weizhang/ E-mail: weizhang@cs.umn.edu
RESEARCH INTERESTS	Computational Biology, Machine Learning, Data Mining, and Biostatistics	
EDUCATION	University of Minnesota Twin Cities, MN, USA Ph.D., Computer Science and Engineering Dec 2015 <ul style="list-style-type: none">Thesis: Computational Analysis of Transcript Interactions and Variants in CancerCommittee: Rui Kuang, Ph.D. (Advisor), Baolin Wu, Ph.D. (Advisor), Jeongsik Yong, Ph.D., Vipin Kumar, Ph.D., Chad L. Myers, Ph.D. M.S., Computer Science and Engineering May 2011 <ul style="list-style-type: none">Advisor: Rui Kuang, Ph.D., Associate Professor Hebei University of Technology, Tianjin, China B.S., Computer Science July 2009 Winona State University, MN, USA <i>Exchange Student</i> , Computer Science Aug 2006 – May 2009 <ul style="list-style-type: none">Advisor: Mingrui Zhang, Ph.D., Professor	
RESEARCH EXPERIENCE	University of Minnesota Twin Cities, MN, USA Research Associate Jan 2016 – present <ul style="list-style-type: none">Developing machine learning and statistical models to analyze biological data and healthcare data. Research Assistant Aug 2009 – Dec 2015 Advisor: Rui Kuang, Ph.D., Associate Professor, Department of Computer Science and Engineering <ul style="list-style-type: none">Developed and implemented a network-based method for RNA-Seq-based transcript quantification with relevant prior knowledge, protein domain-domain interactions.Developed and implemented a network-based Cox regression model to explore the co-expression or functional relation among high-dimensional gene expression features for survival analysis.Developed and implemented multiple signed network propagation algorithms and regularization frame-works for detecting differential gene expressions and DNA copy number variations.Implemented a network-based computational approach rcNet and its webtool PheGenEx to predict associations between disease phenotypes and gene sets. Research Assistant Sept 2010 – Aug 2015 Advisor: Baolin Wu, Ph.D., Associate Professor, Division of Biostatistics, School of Public Health <ul style="list-style-type: none">Developed and implemented a flexible and efficient Gaussian mixture model combined with multidimensional scaling to cluster 16S rRNA sequences for operational taxonomic units prediction.Developed and implemented a lasso penalized Poisson regression model to analyze next generation sequencing data for disease outcome prediction and biomarker discovery. Research Assistant Jan 2013 – Aug 2015 Advisor: Jeongsik Yong, Ph.D., Assistant Professor, Biochemistry, Molecular Biology, and Biophysics <ul style="list-style-type: none">Perform computational analysis to identify alternative 3'-UTR and intronic polyadenylation on mouse embryonic fibroblast cell lines and TCGA data to understand changes of molecular features on dysregulated activation of mammalian target of rapamycin (mTOR). Winona State University, MN, USA Research Assistant Aug 2008 – June 2009 Advisor: Mingrui Zhang, Ph.D., Professor, Department of Computer Science <ul style="list-style-type: none">Developed a fuzzy c-means clustering algorithm which uses a Pearson correlation in its distance metrics and applied it to microarray gene expression data.	
INTERNSHIP	Millennium: The Takeda Oncology Company, Cambridge, MA, USA Research Intern May 2014 – Aug 2014 Advisor: Bin Li, Ph.D., Associate Director, Department of Mathematical and Computational Modeling in Translational Medicine <ul style="list-style-type: none">Developed an integrative model to learn the information from both gene expression and mutation data for drug sensitivity prediction.	

PUBLICATION

1. Lining Liang, Hao Sun, **Wei Zhang**, Mengdan Zhang, Xiao Yang, Rui Kuang and Hui Zheng. "Meta-Analysis of EMT Datasets Reveals Different Types of EMT." *PLoS One*, 2016.
2. **Wei Zhang**, Jae-Woong Chang, Lilong Lin, Kay Minn, Baolin Wu, Jeremy Chien, Jeongsik Yong, Hui Zheng and Rui Kuang. "Network-based Isoform Quantification with RNA-Seq Data for Cancer Transcriptome Analysis." *PLoS Computational Biology*, 2015. doi:10.1371/journal.pcbi.1004465
3. Jae-Woong Chang*, **Wei Zhang***, Hsin-Sung Yeh, Ebbing de Jong, Semo Jun, Kwan-Hyun Kim, Sun Sik Bae, Kenneth Beckman, TaeHyun Hwang, Kye-Seong Kim, Do-Hyung Kim, Rui Kuang and Jeongsik Yong. "mRNA 3'UTR Shortening is a Molecular Signature of mTORC1 Activation." *Nature Communications*, 2015. (*Joint first authors) doi:10.1038/ncomms8218
4. **Wei Zhang**, Takayo Ota, Viji Shridhar, Jerermy Chien, Baolin Wu and Rui Kuang. "Network-based Survival Analysis Reveals Subnetwork Signatures for Predicting Outcomes of Ovarian Cancer Treatment." *PLoS Computational Biology*, Mar 2013. doi:10.1371/journal.pcbi.1002975
5. Jiaqi Yang, **Wei Zhang** and Baolin Wu. "A Note on Statistical Method for Genotype Calling of High-throughput Single Nucleotide Polymorphism Arrays." *Journal of Applied Statistics*, Mar 2013. doi:10.1080/02664763.2013.785499
6. **Wei Zhang**, Nicholas Johnson, Baolin Wu and Rui Kuang. "Signed Network Propagation for Detecting Differential Gene Expressions and DNA Copy Number Variations." *ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM BCB)*, Oct 2012.
7. TaeHyun Hwang, **Wei Zhang**, Maoqiang Xie, Jinfeng Liu and Rui Kuang. "Inferring Disease and Gene Set Associations with Rank Coherence in Networks." *Bioinformatics*, 27(19):2692–2699, 2011.
8. **Wei Zhang**, Baryun Hwang, Baolin Wu and Rui Kuang. "Network Propagation Models for Gene Selection." *Proc. of IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, Nov 2010.
9. Mingrui Zhang, **Wei Zhang**, Hugues Sicotte and Ping Yang. "A New Validity Measure for a Correlation-Based Fuzzy C-means Clustering Algorithm." *Conference Proc. IEEE Engineering in Medicine and Biology Society (EMBC)*, Sept 2009.

PAPERS IN PREPARATION

1. **Wei Zhang**, Jeremy Chien, Jeongsik Yong and Rui Kuang. "Network-based Machine Learning and Graph Algorithms for Precision Oncology." (Under review)
2. Jae-Woong Chang*, **Wei Zhang***, Hsin-Sung Yeh, Ebbing de Jong, Chengguo Yao, Jung-Hee Lee, Hyun-Seo Lee, LeeAnn Higgins, Todd Markowski, Luke Erber, Yue Chen, TaeHyun Hwang, HoJin You, Yongsheng Shi, Timothy Griffin, Rui Kuang and Jeongsik Yong. "mTOR-coordinated Genome-wide Intronic Polyadenylation Reveals Multifaceted Gene Features". (*Joint first authors)
3. Raphael Petegrosso*, **Wei Zhang***, Zhuliu Li, Yousef Saad and Rui Kuang. "Low-rank Label Propagation for Semi-supervised Learning with 100 Millions Samples." (*Joint first authors)
4. **Wei Zhang**, Jae-Woong Chang, Hsin-Sung Yeh, Yue Chen, Rui Kuang and Jeongsik Yong. "Intronic Polyadenylation-mediated Reprogramming of Protein Interaction Networks in Cancers."
5. **Wei Zhang**, Raphael Petegrosso, Jae-Woong Chang, Jeremy R Chien, Jeongsik Yong and Rui Kuang. "Comparative analysis of Isoform-level Gene Expression Estimated by NanoString, RNA-Seq, Exon-Array and qRT-PCR Platforms."
6. Jae-Woong Chang*, **Wei Zhang***, Yongsheng Shi, Rui Kuang and Jeongsik Yong. "A New Integrative Model for Identifying Alternative Polyadenylation from PAS-Seq and RNA-Seq." (*Joint first authors)
7. **Wei Zhang** Rui Kuang, and Baolin Wu. "Model Based Clustering Analysis of 16S rRNA Sequence."

PRESENTATIONS

1. "Computational Analysis of Transcript Interactions and Variants in Cancer." University of Maine, Orono, Oct 2016
2. "Computational Analysis of Transcript Interactions and Variants in Cancer." Duke University, Durham, July 2016
3. "Computational Analysis of Transcript Interactions and Variants in Cancer." Broad Institute, Cambridge, Jan 2016
4. "Network-based Isoform Quantification with RNA-Seq Data for Cancer Transcriptome Analysis."

ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM BCB), Atlanta, Sept 2015

5. “Signed Network Propagation for Detecting Differential Gene Expressions and DNA Copy Number Variations.” *ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM BCB)*, Orlando, Oct 2012

6. “Network-based Survival Analysis on Ovarian Cancer.” *Intelligent Systems for Molecular Biology (ISMB)*, Long Beach, July 2012.

TEACHING
EXPERIENCE

University of Minnesota Twin Cities, MN, USA

Teaching Assistant

Jan 2011 – Dec 2015

CSci 5461 Functional Genomics, Systems Biology and Bioinformatics (Spring 2013/Spring 2011): This is an interdisciplinary graduate level course. The course covers an introduction to various types of functional genomic data and current computational/statistical methods used for analyzing the data to answer questions in functional genomics, systems biology and their applications in disease research. Enrolled students have different backgrounds including biology, biochemistry, and computer science. I work closely with the instructor in this course, giving lecture on Matlab tutorial, planning homework, exercises, and final project.

CSci 8980 Emerging Bioinformatics Models for Next Generation Sequencing and Biological Network Analysis (Fall 2012/Fall 2015): This is an interdisciplinary graduate level course. The lectures in the course cover computational techniques and software tools for analyzing next generation sequencing data and biological networks. Enrolled students have different backgrounds including biology, and computer science. I helped instructor to prepare homework, exercises, and final project, answered the questions from students, and giving lecture and Matlab tutorial.

CSci 5481 Computational Techniques in Genomics (Fall 2011): This is an interdisciplinary graduate level course. This course provides a comprehensive introduction to the fundamental algorithms and mathematical models in modern computational molecular biology. Most enrolled students are computer science background or biology student with programming skill. I helped instructor to prepare homework, exercises, and final project, answered the questions, and giving lecture and Matlab tutorial.

CURRENT
PROJECTS

1. Developing artificial intelligence for scalable and personalized intervention solution for chronic diseases (Nov 2015 – present)

Summary: In the first stage, we developed a physiology model to simulate virtual patients for evaluating and self-training reinforcement learning model, as well as improving understanding of the physiology behind the progression of chronic disease. Then the model will apply to hundreds of firefighters base on their past physiological history and intervention results to continuously identify optimal recommendations for intervention.

2. Genome-wide analysis to identify alternative 3'UTR and intronic polyadenylation (Jan 2014 – present)

Collaborators: Prof. Jeongsik Yong (PI, Biochemistry, Molecular Biology, and Biophysics).

Summary: Perform genome-wide analysis to determine cancer-enriched intronic and 3'UTR alternative polyadenylation (APA) in eleven TCGA cancer types by referring the catalog of mTOR-dependent APA obtained from cancer cell lines, and to characterize their functions in the modulation of cellular signaling pathways and in cell proliferation. In addition, an integrative model is built up by integrating RNA-Seq and PAS-Seq data to improve the accuracy of predicting the APA with RNA-Seq data only.

3. Comparative analysis of isoform-level gene expression estimated by NanoString data and other platforms (Oct 2014 – present)

Summary: In this project, we estimated isoform and gene expression from 50 cancer cell lines by using our custom-designed NanoString, RNA-Seq, Microarray, Exon-Array and qRT-PCR platforms. We compare the expressions and fold-change of expressions estimated by different platforms. In addition, tissue-specific isoforms and genes are identified.

4. Low-rank label propagation for semi-supervised learning (January 2016 – present)

Summary: In this project, we adopt the idea of two low-rank label propagation algorithms, GLNP (Global Linear Neighborhood Propagation) and Kernel Nyström Approximation, and implemented the parallelized version of the two algorithms for big data label propagation. The parallel algorithms are scalable for learning with tens of million or more data points. The whole workflow is useful for solving biological problems, e.g., homology detection and SNP prediction.

5. Drug sensitivity prediction with genome and transcriptome sequencing data (Jan 2016 – present)
Collaborators: Prof. Taehyun Hwang (Department of Clinical Science, University of Texas Southwestern Medical Center).

Summary: In this project, we develop a workflow for drug sensitivity prediction by integrating transcriptomic and genomic profiles. RNA-Seq gene expression, somatic mutation, and copy number variations information extract from more than 100 lung cancer cell lines and dose-response information (ED_{50}) of more than 250 compounds are supplied. Six different prediction methods are applied in the analysis.

COMPLETE
PROJECTS

1. Network-based transcript quantification with RNA-Seq data (Mar 2013 – July 2015)

Advisor: Prof. Rui Kuang,

Collaborators: Dr. Hui Zheng (Guangzhou Institutes of Biomedicine and Health of Chinese Academy of Sciences), Prof. Jeremy Chien (University of Kansas Cancer Center), Prof. Baolin Wu (Division of Biostatistics, University of Minnesota).

Summary: In this project, we propose the idea of integrating protein domain-domain interaction information with short read alignment for transcript abundance estimation. We developed a Network-based model for RNA-Seq-based Transcript Quantification (Net-RSTQ) based a transcript network constructed from protein domain-domain interaction databases and an optimization algorithm to solve the model. qRT-PCR experiments and classification results on the RNA-Seq data of ovarian cancer, breast cancer and lung cancer in The Cancer Genome Atlas (TCGA) suggest that the transcript network information is indeed useful for RNA-Seq-based transcriptome analysis. This work has been published in **PLoS Computational Biology**.

2. mTOR reprograms alternative polyadenylation factors for transcriptome-wide control of 3' UTR length (Jan 2013 – Feb 2015)

Advisor: Prof. Rui Kuang,

Collaborators: Prof. Jeongsik Yong (PI, Biochemistry, Molecular Biology, and Biophysics, University of Minnesota).

Summary: In this project, we identifies 3' UTR shortening as a new cellular target of mTOR pathway and provides a novel mechanistic insight of how the mTOR pathway modulates gene expression by altering post-transcriptional gene regulation mechanisms. The mammalian target of rapamycin (mTOR) pathway plays a crucial role in cell proliferation/growth and connects to pathogenesis of many diseases. To better understand the function of mTOR at high-resolution, we investigated the mTOR-activated transcriptome features and discovered transcriptome-wide 3' untranslated region (UTR) shortening in mRNA. Using a genetic or chemical modulation of mTOR activity in cells or in terminally differentiated mouse skeletal muscle, we demonstrate that 3' UTR shortening is dependent on mTOR activity. This work has been published in **Nature Communications**.

3. Model based clustering analysis of 16S rRNA sequence (Nov 2012 – Jan 2015)

Advisors: Prof. Baolin Wu (Division of Biostatistics, University of Minnesota)

Summary: In this project, we propose a flexible and efficient gaussian mixture model combined with multidimensional scaling to cluster 16S rRNA sequences for operational taxonomic units (OTUs) prediction. Through applications to public datasets and numerical simulations, we demonstrate that the proposed method performs better than existing methods.

4. Network-based survival analysis on ovarian cancer (July 2011 – Dec 2012)

Advisor: Prof. Rui Kuang,

Collaborators: Prof. Jeremy Chien (University of Kansas Cancer Center), Prof. Baolin Wu (Division of Biostatistics, University of Minnesota).

Summary: In this project, we propose a general network-based survival model, named Net-Cox, for survival analysis on high-dimensional data constrained by a network describing the relation among features. The Cox proportional hazard model is widely used to predict or correlate gene expressions with time to an event of interest (outcome) in cancer genomics. Net-Cox is the first model to integrate gene network information into the Cox proportional hazard model to explore the co-expression or functional relation among high-dimensional genomic features in the gene network. We studied resistance to chemotherapy in ovarian cancer using Net-Cox with a focus on the survival and recurrence in the treatment of ovarian carcinoma and confirmed with experiments that networks representing gene co-expression or functional relations could be used to improve the accuracy and the robustness of survival prediction of outcome in ovarian cancer treatment. This work has been published in **PLoS Computational Biology**.

5. Signed network propagation models for cancer outcome prediction and biomarker discovery (Oct 2009 – May 2011)

Advisor: Prof. Rui Kuang.

Summary: In this project, we developed several network propagation algorithms for biomarker discovery from high-dimensional large-scale gene expression data and copy number variation data. We showed that these new algorithms integrating gene co-expression information into biomarker discovery as a network improved the robustness of biomarker discovery from high-dimensional microarray data such that the biomarkers identified from independent studies are more consistent. We also derived a new network propagation algorithm for signed network, with which the network propagation framework can be applied to gene networks with both positive and negative co-relations. This algorithm further enhanced the application of network propagation to biomarker discovery. The results of these techniques have been published in **ACM BCB 2012**, and **GENSIPS 2010**.

AWARDS

- Best Poster Award, The 6th Annual Biomedical Informatics and Computational Biology (BICB) Research Symposium, 2014
- CNB MAC 2015 Doctoral Student Travel Grant, 2015
- ACM BCB 2012 Doctoral Student Travel Grant, 2012
- 4th Place of Annal NCS-MAA Mathematical Contest, 2007

PROFESSIONAL SERVICES

Reviwer: Bioinformatics, PLoS One, BMC Bioinformatics, BMC Genomics, BMC Medical Genomics. External reviewer: Nucleic Acids Research, ISMB, RECOMB, KDD, ICDM, SDM, NIPS.

COMPUTER SKILLS

MATLAB, R, Java, C, UNIX shell scripting, Prolog, L^AT_EX, Inkscape, and Cytoscape. Knowledge of leading next-gen sequencing tools: SAMtools, GATK, BWA, Bowtie, TopHat, HISAT, Kallisto and Cufflinks.

REFERENCES

Rui Kuang, Ph.D.

Associate Professor
Department of Computer Science and Engineering,
University of Minnesota Twin Cities, MN 55455
Phone: (612)624-7820
E-mail: *send.Kuang.D76C8C5F61@interfolio.com*

Baolin Wu, Ph.D.

Associate Professor
Division of Biostatistics, School of Public Health
University of Minnesota Twin Cities, MN 55455
Phone: (612)624-0647
E-mail: *send.Wu.B950B77546@interfolio.com*

Vipin Kumar, Ph.D.

Regents Professor
Department of Computer Science and Engineering,
University of Minnesota Twin Cities, MN 55455
Phone: (612)624-8023
E-mail: *send.Kumar.B69D23F217@interfolio.com*

Jeongsik Yong, Ph.D.

Assistant Professor
Department of Biochemistry, Molecular Biology, and Biophysics
University of Minnesota Twin Cities, MN 55455
Phone: (612)626-2420
E-mail: *send.Yong.60D61AE707@interfolio.com*

Bin Li, Ph.D.

Associate Director, Computational Biologist
Department of Mathematical and Computational Modeling in Translational Medicine
Millennium: The Takeda Oncology Company, Cambridge, MA 02139
Phone: (617)679-7041
E-mail: *send.Li.74FF50919E@interfolio.com*