

2017-09-25

Guoshuai Cai, PhD

Department of Environmental Health Sciences
Arnold School of Public Health
University of South Carolina
Phone: 803-777-4120

921 Assembly Street
Public Health Research Center 401C
Columbia, SC, USA
E-mail: GCAI@mailbox.sc.edu

EDUCATION

- Aug 2013 **PhD, Biomathematics and Biostatistics**
Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston and MD Anderson Cancer Center, Houston, TX, USA
Mentor: Shoudan Liang, PhD
Dissertation: *RNA-Sequencing applications: Gene Expression Quantification and Methylator Phenotype Identification*
- Jun 2009 **Master of Science, Microbiology**
State Key Laboratory of Virology
College of Life Sciences, Wuhan University, Wuhan, China
- Jun 2006 **Bachelor of Science, Biotechnology**
College of Life Sciences, Wuhan University, Wuhan, China

PROFESSIONAL EXPERIENCE

- Aug 2017 **Assistant Professor**
-
Now Department of Environmental Health Sciences, Arnold School of Public Health,
University of South Carolina, Columbia, SC, USA
- Nov 2014 **Postdoctoral Fellow**
-
Aug 2017 Department of Molecular and Systems Biology, The Geisel School of Medicine at
Dartmouth, Hanover, NH, USA
Mentor: Michael Whitfield, PhD
Project I: *Methodology development for RNA-seq differential expression analysis with length bias adjustment.*
Project II: *Identification and selection of transcriptome markers predicting scleroderma disease severity.*
Project III: *Methodology to infer intercontinental ancestry using genetic data*

2017-09-25

Project IV: *Methodology to infer subset structures using transcriptome expression profiles*

Project V: *Population stratification identifies gene expression predictors of survival outcomes in lung adenocarcinoma for both Caucasian and Asian patients.*

Oct 2013 **Postdoctoral Research Associate**

- Department of Clinical Cancer Prevention, MD Anderson Cancer Center,
Houston, TX, USA

Oct 2014 Mentor: Qiang Shen, MD, PhD

Project I: *Study of mechanisms in cell of new anticancer drug compounds by high-throughput data analysis.*

Project II: *Identification of genes driving breast cancer metastasis.*

Project III: *A computing platform for comprehensive statistical analysis of large-scale and high-dimensional biomedical data.*

Dec 2009 **Graduate Research Assistant**

- Department of Bioinformatics and Biostatistics, The University of Texas Health
Jun 2013 Science Center at Houston MD Anderson Cancer Center, Houston, TX, USA

Mentor: Shoudan Liang, PhD

Project I: *Modeling the overdispersion in RNA-seq with its dependence on sequencing depth.*

Project II: *Identification of a new bias resource in sequencing technology and a statistical method for correction.*

Project III: *Identification of a new CpG island methylator phenotype in breast cancer by integrating methylation and mRNA expression.*

Dec 2006 **Graduate Research Assistant**

- Department of Neuroscience, University of Minnesota, Minneapolis, MN, USA

Nov 2008 Supervisor: Tongbin Li, PhD

Project: *Development of miRecords, an integrated resource for microRNA-target interactions.*

RESEARCH INTERESTS

Methodology and Application in:

Sequencing data analyses
Disease prediction and marker selection
Time series data analyses

Integrative genomics
Structure inference
Cancer and scleroderma

PUBLICATIONS

Published or in press:

1. Zhao Y, Varn F, **Cai G**, Xiao F, Amos CI, and Cheng C. A P53-deficiency gene signature predicts recurrence risk of patients with early stage lung adenocarcinoma. *Cancer Epidemiology, Biomarkers & Prevention*. In press.
2. **Cai G**, Zheng X, Liang S, Xiao F. Local sequence and sequencing depth dependent accuracy of RNA-seq reads. *BMC Bioinformatics*. 2017 Aug 9;18(1):364.
3. **Cai G**, Xiao F, Cheng C, Li Y, Amos CI, Whitfield ML. Population stratification identifies gene expression predictors of survival outcomes in lung adenocarcinoma for both Caucasian and Asian patients. *Plos One*. 2017 Apr 20;12(4):e0175850.
4. Xiao F, **Cai G (co-first author)**, Zhang H. Segregation Analysis Suggests That a Genetic Reason May Contribute to "the Dress" Colour Perception. *Plos One*. 2016 Oct 21;11(10):e0165095.
5. Li Y, Byun J, **Cai G (co-first author)**, Xiao X, Dennis J, Easton D, Gorlov I, Seldin M, Amos CI. FastPop: a rapid principle component derived method to infer intercontinental ancestry using genetic data. *BMC Bioinformatics*. 2016 Mar 9;17:122.
6. Bu Y, **Cai G**, Shen Y, Huang C, Cao Y, Liao D, Cao D. Targeting NF- κ B RelA/p65 Phosphorylation Overcomes RITA resistance. *Cancer Letters*. 2016 Oct 6;383(2):261-271.
7. Zhou X, Ren Y, Kong L, **Cai G (co-first author)**, Sun S, Song W, Wang Y, Jin R, Qi L, Mei M, Wang X, Kang C, Li M, Zhang L. Targeting EZH2 regulates tumor growth and apoptosis through modulating mitochondria dependent cell-death pathway in HNSCC. *Oncotarget*. 2015 Oct 20;6(32):33720-32.
8. Zhou X, Liu S, **Cai G (co-first author)**, Kong L, Zhang T, Ren Y, Wu Y, Mei M, L, and Wang X. Long Non Coding RNA MALAT1 Promotes Tumor Growth and Metastasis by inducing Epithelial-Mesenchymal Transition in Oral Squamous Cell Carcinoma. *Scientific report*. 2015 Nov 2; 5:15972.
9. Li H, Tong P, Gallegos J, Dimmer E, **Cai G**, Molldrem J.J, Liang S. PAND: a distribution to identify functional linkage from networks with preferential attachment property. *Plos One*. 2015; 10(7).
10. Lee J, Ji Y, Liang S, **Cai G**, Müller P. Bayesian Hierarchical Model for Differential Gene Expression Using RNA-seq Data. *Statistics in Biosciences*. 2015; 7(1).
11. Tang SH, Wang X, Shen Q, Yang X, Yu C, Cai C, **Cai G**, Xu L, Meng X, and Zou F. Mitochondrial Ca²⁺ uniporter is critical for store-operated Ca²⁺ entry-dependent breast cancer cell migration. *Biochem Biophys Res Commun*. 2015; 458(1):186-93.
12. Chen H, Yang Z, Ding C, Xiong A, Wild C, Wang L, Ye N, **Cai G**, Flores RM, Ding Y, Shen Q, Zhou J. Eur J Med Chem. Discovery of potent anticancer agent HJC0416, an orally bioavailable small molecule inhibitor of signal transducer and activator of transcription 3 (STAT3). *Eur J Med Chem*. 2014; 82:195-203.
13. Xiao F, Ma J, **Cai G**, Amos CI. Natural and Orthogonal interaction framework for modeling gene-gene interactions applied to cutaneous melanoma. *Human Genetics*. 2014; 133(5):559-74.
14. **Cai G**, Li H, Lu Y, Huang X, Lee J, Müller P, Ji Y, Liang S. Accuracy of RNA-Seq and its dependence on sequencing depth. *BMC Bioinformatics*. 2012; 13:S5.
15. Lee J, Ji Y, Liang S, **Cai G**, Müller P. On differential gene expression using RNA-Seq data. *Cancer Inform*. 2011; 10:205-15.
16. Xiao F, Zuo Z, **Cai G**, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res*. 2009; 37:D105-10.

17. Kim Y, Xiao H, Du E, **Cai G**, Lu S, Qi Y. Identification and functional analysis of LsMNPV anti-apoptosis genes. *J Biochem Mol Biol.* 2007; 40(4):571-6.

Under review or in preparation:

18. Franks JM, **Cai G**, and Whitfield ML. Whitfield Feature specific quantile normalization enables cross-platform classification of molecular subtypes using gene expression. *Bioinformatics*, Submitted.
19. Zhang L, Cao W, Majewski T, Yuan Y, Young K, Broom B, **Cai G**, Lu K, Sun S, Czerniak B, Weinstein J, Zhang L. Somatic DNA copy number variants in the peripheral blood of patients with solid tumors. *Nucleic Acids Res.* Submitted.
20. Messemaker TC, Chadli L, **Cai G**, Goelela VS, Boonstra M, Dorjée AL, Andersen SN, Mikkers HM, van t Hof P, Mei H, Distler OMD, Draisma HHM, Li Z, Toes REM, Aarbiou J, Huizinga TW, Whitfield M, DeGroot J, de Vries-Bouwstra JK, Kurreeman Fina. Antisense long non-coding RNAs are deregulated in skin tissue of patients with systemic sclerosis. *Journal of Investigative Dermatology.* Under review.
21. **Cai G**, Xiao F. RNA-seq differential expression detection and its study specific strategy. Under revision.
22. **Cai G**, Qiu P, Verhaak RG, Chen T, Xiao F, Liang S. Expression regulatory methylator phenotype in breast cancer. Under revision.
23. **Cai G**, Xiao F, Liang S. A new bias in RNA-seq. Under revision.
24. Yang Z, **Cai G (co-first author)**, Chen H, Ding C, Zhang Y, Xiong A, Ester C, Wild C, Zhou J, and Shen Q. Orally active small-molecule inhibitor HJC0152 exhibits potent anticancer activity and suppresses the growth of triple-negative breast cancer xenograft tumors via inhibiting STAT3/JAK signaling. Under revision.
25. Zheng X, Lu Y, **Cai G**, Zhang J, Liang S. MTM: An error-correction method for high-throughput sequencing data. Manuscript.
26. **Cai G**, Whitfield ML. Precision weighted mRNA abundance in RNA-seq differential expression analysis. Manuscript.
27. **Cai G**, Franks J, Whitfield M. Subset calibration through concordance of gene expression profiles. Manuscript.
28. **Cai G**, Yang Z, Zhang Y, Jiang S, Price J, Brown P, Shen Q. Role of the AP-1 transcription factor in promoting breast cancer metastasis. Manuscript.
29. **Cai G**, Shen Q. Networks involved in the metastatic phenotype of highly metastasis variants of breast cancer cells. Manuscript.
30. **Cai G**, Shen Q. Comparison of STAT3 inhibitors Niclosamide, HJC0152, HJC0123, HJC0416 and Stattic. Manuscript.
31. **Cai G**, Whitfield M, Improved RNA-seq differential analysis method with length bias adjustment. In preparation.
32. **Cai G**, Franks J, Whitfield M. Identification of a miRNA regulatory circuit for severity of scleroderma skin disease. In preparation.
33. Yang Z, **Cai G**, Zhang Y, Jiang S, Price J, Brown P, Shen Q. Signal transduction pathways and gene sets responsible for an enhanced metastatic breast cancer phenotype. In preparation.

SOFTWARES AND TOOLS

- **miRecords**: an integrated resource for microRNA-target interactions

- **FastPop**: a rapid principle component derived method to infer intercontinental ancestry using genetic data
- **GCAI.bias**: a guided correction approach for sequencing inherent bias
- **RoMA**: RNA-seq analysis on Molecular Abundance.
- **ssGSEA**-seq: single sample gene set enrichment analysis for RNA-seq data
- **gladR**: gene length adjusted differential expression analysis for RNA-seq data
- **BONE**: a biomedical big data comprehensive analysis system

INVITED TALKS

1. Integrated genome-wide analysis of DNA Methylation and RNA Expression. Invited talk, School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, China, Jul 2017.
2. Steps to precise disease subtyping. Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, USA. Feb 2017.
3. Steps to precise disease subtyping. Department of Environmental Health Science, University of South Carolina, Columbia, SC, USA. Feb 2017.
4. Differential Expression and Methylation-expression Integration Analyses in RNA-seq. University of South Carolina Statistics Department Colloquium, Columbia, SC, USA. Mar 2016.
5. Problems and Methods of RNA-seq data differential expression analyses. Genomics Working Group. Hanover, NH, USA. Apr 2015.

CONFERENCE PRESENTATIONS AND ABSTRACTS

1. Transcriptome sequencing reveals genetic polymorphisms associated with SSc gene expression subtypes. *Poster Presentation*. The 15th international workshop on scleroderma research, Pittsburgh, PA, USA. Aug 2017.
2. Novel machine learning classifier accurately predicts intrinsic molecular subsets for patients with systemic sclerosis. *Abstract*. The 15th international workshop on scleroderma research, Pittsburgh, PA, USA. Aug 2017.
3. In vitro skin models mimic fibrogenic signatures of systemic sclerosis. *Abstract*. The 15th international workshop on scleroderma research, Pittsburgh, PA, USA. Aug 2017.
4. Intrinsic subsets are conserved across multiple organs in systemic sclerosis. *Abstract*. The 15th international workshop on scleroderma research, Pittsburgh, PA, USA. Aug 2017.
5. Reprogramming glucose metabolism and energy production with a small molecule HJC0152 suppresses breast cancer development and progression to metastasis. *Abstract*. American Association for Cancer Research (AACR) Annual Meeting, New Orleans, Louisiana, Apr 2016.
6. Targeting STAT3 with novel small molecule inhibitors to sensitize breast cancer cells to radiation therapy. *Abstract*. The Thirty-Seventh Annual CTSC-AACR San Antonio Breast Cancer Symposium. San Antonio, TX. Dec, 2014.

2017-09-25

7. Using expression data to define patient specific predictors for survival outcomes in lung adenocarcinoma. *Poster Presentation*. American Society of Human Genetics Annual Meeting (ASHG), Baltimore, MD, USA. Oct 2015.
8. Networks Involved in the Metastatic Phenotype of Highly Metastasis Variants of Breast Cancer Cells. *Presentation*. 28th Society of Chinese Bioscientists in America (SCBA), Houston, TX, USA. Jun 2014.
9. Accuracy of RNA-seq and its dependence on sequencing depth. *Poster presentation*. Advances in Genome Biology and Technology (AGBT), Miami, FL, USA. Feb 2012.
10. Accuracy of RNA-Seq and its dependence on sequencing depth. *Presentation*. Biotechnology and Bioinformatics Symposium (BIOT), Houston, TX, USA. Jun 2011.

SHORT COURSES/EDUCATIONAL WORKSHOPS

1. Educational Workshop on Genetic Epidemiology: Fundamentals for Precision Medicine, Cambridge, UK. Sep 2017

OTHER ACTIVITIES

- Reviewer for *Journal of Investigative Dermatology*, *immunobiology*, *Journal of Thoracic Oncology*, *BioData Mining*, *European Conference on Computational Biology 2016*, *Molecular and Cellular Biochemistry*, *Journal of Diabetes Investigation* and *Journal of Cancer Therapy*, *PeerJ*.
- Member of the International Genetic Epidemiology Society (IGES)
- Member of the American Society of Human Genetics (ASHG)
- Member of the American Statistical Association (ASA)