REQUIRED CURRICULUM VITAE FORMAT

PROMOTION AND TENURE

NAME: Zhi Wei

|  |  |
| --- | --- |
| Home Address: | Faculty Address: |
| 112 Lowell Ct Apt 10 | Faculty Rank: Assistant Professor |
| Princeton, NJ 08540 | Primary Department: Computer Science |
| Home Telephone: 215-900-2308 | Office Telephone: 973-642-4497 |
| Home e-mail address: zhiwei@njit.edu | Campus e-mail address: zhiwei@njit.edu |
|  | Joint Appointment(s): |

**I. EDUCATION**

# **Formal**

|  |  |  |  |
| --- | --- | --- | --- |
| Degree | Area | Year Granted | College/University |
| Ph.D. | Bioinformatics | 2008 | University of Pennsylvania |
| M.S. | Computer Science | 2004 | Rutgers University - New Brunswick |
| M.E. | Computer Science | 2002 | Wuhan University, China |
| B.S. | Computer Science | 2000 | Wuhan University, China |

# **Other Education/Special Courses**

|  |  |  |  |
| --- | --- | --- | --- |
| Course | Location | Description | Date |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**II. RESEARCH INTERESTS** (Provide clear statements of current research interests, significant accomplishments, and plans for future research in the CV. Additional primary information is to be submitted in the full dossier.)

My research interests are in the following areas:

* Bioinformatics
* Statistical genetics
* Large-scale multiple testing
* High-dimensional data analysis
* Empirical Bayes methods
* Machine learning methods
* Algorithms

My work focuses on developing statistical and machine learning methods as well as computer algorithms for the analysis of high dimensional data. Most of my methods are motivated by cutting-edge biology and genetics problems. After joining NJIT, I have developed quite a few popular computational methods to solve those challenging problems for analysis of various biological data, including microarray gene expression data, SNParray data, DNA-sequence data, and the recent next-generation sequencing data. This methodology research work has resulted in publications in the most prestigious computational journals and computer science conference proceedings including *Bioinformatics*, the No.1 bioinformatics journal, *Journal of the American Statistical Association*, one of the top four well-recognized statistics journals, *PLoS Genetics*, a top genetics journal, *Nucleic Acids Research*, a high-impact biology journal publishing highly-selective informatics methods/tools, and *NIPS*, a top-tier machine learning conference.

I have established extensive collaborations with well-known biologists, geneticists, and physicians from big biology laboratory, cancer institute, hospital, genomics center, and international genetics consortium. This collaborative work has led to publications in reputable journals, such as *Nature*, *Pigment Cell Melanoma Research* and *Physiological Genomics*.

In the future, I will focus on developing computational methods for next-generation sequencing (NGS) data analysis. NGS data analysis imposes unprecedented computational challenge and statistics challenge. It also has a higher prerequisite for biology background knowledge to solve those challenges. All these fit exactly well with my expertise. I believe my future work will make more significant contribution to the fields.

**III. EXPERIENCE** (summarize in the CV, additional primary information to be submitted in the full dossier)

## **Academic Appointments**

(include joint appointments and formal post doctoral positions)

|  |  |  |  |
| --- | --- | --- | --- |
| Title | College/University | Date | Tenure (Date) |
| Assistant Professor | New Jersey Institute of Technology | 2008 - present |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

### Non-academic Employment

|  |  |  |
| --- | --- | --- |
| Title | Employer | Date |
|  |  |  |
|  |  |  |
|  |  |  |

### Consulting

|  |  |  |
| --- | --- | --- |
| Organization and/or nature of work | Employer | Date |
|  |  |  |
|  |  |  |
|  |  |  |

FOR REFEREED PUBLICATIONS, INCLUDE REFEREEING PROCEDURES.

**IV. TEACHING ACTIVITIES** (summarize in the CV, additional primary information to be submitted in the full dossier; provide mean ratings for Question 13 from the student evaluations for all courses taught since date of hire)

1. Classroom Evaluations, (submit results per course)

Scores listed below are out of 4.0 and related to instructor’s teaching

Fall, 2008

BNFO 615-101, BIOINFORMATICS-CENTRIC DATA STR, avg rating of Q13: 2.61

Spring, 2009

CS 103-2, COMPUT SCI-BUSINESS PROB, avg rating of Q13: 1.54

Fall, 2009

BNFO 615-101, DATA ANALYSIS IN BIOINFORMATICS, avg rating of Q13: 3.23

Spring, 2010

CS 698-110, ST:HIGH DIMENSIONAL DATA ANALY, avg rating of Q13: 3.57

Fall, 2010

BNFO 615-101, DATA ANALYSIS IN BIOINFORMATICS, avg rating of Q13: 3.36

Spring, 2011

CS 698-110, ST: MACHINE LEARNING, avg rating of Q13: 2.85

Fall, 2011

BNFO 615-101, DATA ANALYSIS IN BIOINFORMATICS, avg rating of Q13: 3.33

Mean score: 2.9/4.0

Mean score excluding first year: 3.3/4.0

New Course(s) Developed

|  |  |  |
| --- | --- | --- |
| Course | Title | Dates Offered |
| BNFO 620 | Genomic Data Analysis | Spring 2012 |
| BNFO 340 | Data Analysis for Bioinformatics | Spring 2010 |
|  |  |  |

**BNFO 620:** With the rapid development of high throughput bio-technology, next-generation sequencing (NGS) becomes very hot in the past few years and is expected to revolutionize many fields in biology. Genomic data analysis, in particular NGS data, is touted as one of the focuses in the next wave of Bioinformatics field. Large-scale genomic data analysis skills and experience are highly desired by both industry and academia employers. It is beneficial for Bioinformatics majors to learn these analytical skills and gain practical experience from class. So I develop this course. This course has high requirement for computing resource and storage. I coordinate with UCS (University Computing Systems) staff to ensure that we have necessary infrastructure resource for this course.

**BNFO 340:** Bioinformatics becomes more and more popular, which makes it critical to establish quantitative view and master analytical approaches for bioinformatics and biology majors. We had BNFO 135, 136 and 240, all of which are basic courses for bioinformatics and biology majors, but no advanced courses. To fill this gap, I develop this course to provide advanced training in this area.

1. Manual(s) Developed and Course or Laboratory Notes

Course notes developed for

* BNFO 615 Data Analysis in Bioinformatics
* BNFO 620 Genomic Data Analysis

1. Teaching Related Publications
2. Research Supervision

Thesis Advisor

Wei Wang, Dept of Computer Science, Ph.D. in progress

Xiao Ling, Bioinformatics Program, M.S. in progress

Yuanpeng Lu, Bioinformatics Program, M.S. in progress

Jharna Miya, Bioinformatics Program, M.S. in progress

Thesis Committee

Zhiying Qiu Dept of Dept of Math. Sciences, Ph.D. 2014

Sugata Banerji, Dept of Computer Science, Ph.D. 2013

Chandralekha De, Dept of Computer Science, Ph.D. 2012

Dibyendu Chakrabarti, Dept of Computer Science, Ph.D. 2012

Yanzhi Bai, Dept of Computer Science, Ph.D. 2010

Satish Chikkagoudar, Dept of Computer Science, Ph.D. 2009

Sinan Ramazanoglu, Bioinformatics Program, M.S., 2012

Drew Roberts, Bioinformatics Program, M.S., 2012

Neha Singh, Bioinformatics Program, M.S., 2012

Seif Shahidain, Bioinformatics Program, M.S., 2011

Meera Prasad, Bioinformatics Program, M.S., 2011

Paras Garg, Bioinformatics Program, M.S., 2009

1. Other Pertinent Materials (e.g. Course Supervision)

**V.** **SCHOLARLY ACTIVITIES** (provide in chronological order, newest publications first) List author’s names in the sequence as they appear in print. Underline primary author. (Summarize in the CV, additional primary information to be submitted in the full dossier Include copies of all publications including book chapters, published articles, articles accepted for publication and published conference papers.)

**A. 1. Published Books** (reviews/adoption lists may be submitted)

##### N/A

**2. Published Book Chapters**

1. **Zhi Wei**, Minturn, J., Rappaport, E., Brodeur, G., and Hongzhe Li, “Network-based Analysis of Multivariate Gene Expression Data”, ***Statistical Methods for Microarray Data Analysis***, Yakovle A, Klebanov L and Gaile G (Ed.), Springer, New York, in press.
2. **Zhi Wei**, “Hidden Markov Models for Controlling False Discovery Rate in Genome-Wide Association Analysis”, ***Next Generation Microarray Bioinformatics***, Junbai Wang (Ed.), Springer, New York, 2012.

**B. 1. Published Refereed Journal Papers** (\* corresponding author, # students, the number of citations is based on google scholar as of Nov 2012)

**At NJIT**

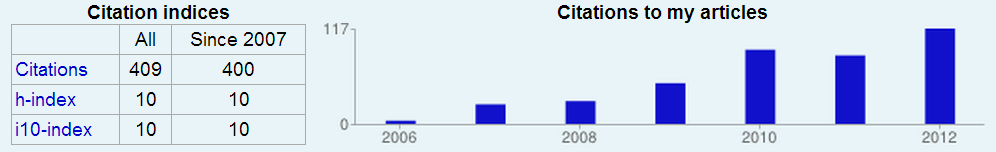
1. Z. Zhao, W. Wang#, and **Zhi Wei\***. An Empirical Bayes Testing Procedure for Detecting Variants in Analysis of Next Generation Sequencing Data, ***Annals of Applied Statistics***, invited revision, Peer-reviewed.
2. Ruifeng Yang, Ying Zheng, Shujing Liu, Michelle Burrows, **Zhi Wei**, George Cotsarelis, Xiaowei Xu, Generation of folliculogenic human epithelial stem cells from induced pluripotent stem cells, ***Nature Biotechnology***, invited revision, Peer-reviewed.
3. Jin Li, Joseph T Glessner, Haitao Zhang, Cuiping Hou, **Zhi Wei**, Jonathan P Bradfield, Frank D Mentch, Yiran Guo, Cecilia Kim, Qianghua Xia, Rosetta M Chiavacci, Kelly A Thomas, Haijun Qiu, Struan FA Grant, Susan L Furth, Hakon Hakonarson, Patrick MA Sleiman, GWAS of Blood Cell Traits Identifies Novel Associated Loci and Epistatic Interactions in Caucasian and African American Children, ***Human Molecular Genetics***, invited revision, Peer-reviewed.
4. Haitao Hu, Martin Nau, Phil Ehrenberg, Agnes Laurence Chenine ,Camila Macedo, Yu Zhou, Z. John Daye, **Zhi Wei**, Maryanne Vahey, Nelson Michael,,Jerome Kim, Mary Marovich, Silvia Ratto-Kim, Distinct gene expression profiles associated with the susceptibility of pathogen-specific CD4+ T cells to HIV-1 infection, ***Blood***, invited revision, Peer-reviewed.
5. Bruce A. Ong, Jin Li, Joseph M. McDonough, **Zhi Wei**, Cecilia Kim, Rosetta Chiavacci, Frank Mench, Jason B. Caboot, Jonathan Sperlgel, Julian L. Allen, Patrick M.A. Sleiman, Hakon Hakonarson, Gene Network Analysis in a Pediatric Cohort Identifies Novel Lung Function Genes, ***European Respiratory Journal***, under review, Peer-reviewed.
6. W. Wang#, W. Hu#, F. Hou#, P. Hu and **Zhi Wei**\*, “SNVerGUI: A desktop tool for variant analysis of next generation sequencing data”, ***Journal of Medical Genetics,*** accepted, Peer-reviewed.
7. Luke Jostins, Stephan Ripke, Rinse K Weersma, Richard H Duerr, Dermot P McGovern, Ken Y Hui, James C Lee, L Philip Schumm, Yashoda Sharma, Carl A Anderson, Jonah Essers, Mitja Mitrovic, Kaida Ning, Isabelle Cleynen, Emilie Theatre, Sarah L Spain, Soumya Raychaudhuri, Philippe Goyette, **Zhi Wei**, et al. “Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease”, ***Nature***, Vol. 491, No. 7422, 2012, pp. 119-124, Peer-reviewed.
8. Renata Pellegrino, Daniele Sunaga, Camila Guindalini, Raquel Martins, Diego Mazzotti, **Zhi Wei**, John Daye, Monica Andersen, and Sergio Tufik, "Peripheral genome-wide gene expression profiles in humans after prolonged wakefulness and sleep recovery", ***Physiological Genomics***, Vol. 44, No. 21, 2012, pp. 1003-1012, Peer-reviewed.
9. Yingbin Ge, Rikka Azuma, Bethsebah Gekonge, Alfonso Lopez-Coral, Min Xiao, Gao Zhang, Xiaowei Xu, Luis J Montaner, **Zhi Wei**, Meenhard Herlyn, Tao Wang, and Russel E Kaufman, “Induction of Metallothionein Expression During Monocyte to Melanoma-Associated Macrophage Differentiation,” ***Frontiers in Biology****,* Vol. 7, No. 4, 2012, pp. 359-367, Peer-reviewed.
10. Z. J. Daye, H. Li, and **Zhi Wei\***, “A Powerful Test for Multiple Rare Variants Association Studies That Incorporates Sequencing Qualities,” ***Nucleic Acids Research****,* Vol. 40, No. 8, 2012, pp. e60:1-12, Peer-reviewed. Citations 1
11. T. Wang, Y. Ge, M. Xiao, A. Lopez-Coral, R. Azuma, R. Somasundaram, G. Zhang, **Zhi Wei**, X. Xu, F. J. Rauscher III, M. Herlyn, and R. E. Kaufman, “Melanoma-Derived Conditioned Media Efficiently Induce the Differentiation of Monocytes to Macrophages That Display a Highly Invasive Gene Signature,” ***Pigment Cell Melanoma Research****,* Vol. 25, No. 4, 2012, pp. 493-505, Peer-reviewed. Citations 2.
12. G. J. Lyon, T. Jiang, R. Van Wijk, W. Wang#, P. M. Bodily, J. Xing, L. Tian, R. J. Robison, M. Clement, Y. Lin, P. Zhang, Y. Liu, B. Moore, J. T. Glessner, J. Elia, F. Reimherr, W. W. van Solinge, M. Yandell, H. Hakonarson, J. Wang, W. E. Johnson, **Zhi Wei**, and K. Wang, “Exome Sequencing and Unrelated Findings in the Context of Complex Disease Research: Ethical and Clinical Implications,” ***Discovery Medicine****,* Vol. 12, No. 62, 2011, pp. 41-55, Peer-reviewed. Citations 13.
13. W. Sun, and **Zhi Wei** (alphabetical order), “Multiple Testing for Pattern Identification, with Applications to Microarray Time-Course Experiments,” ***Journal of the American Statistical Association****,* Vol. 106, No. 493, 2011, pp. 73-88, Peer-reviewed. Citations 5.
14. W. Wang#, **Zhi Wei**, T. W. Lam, and J. Wang, “Next Generation Sequencing Has Lower Sequence Coverage and Poorer Snp-Detection Capability in the Regulatory Regions,” ***Scientific Reports****,* Vol. 1, 2011, pp. 55, Peer-reviewed. Citations 4.
15. U. Roshan, S. Chikkagoudar, **Zhi Wei**, K. Wang, and H. Hakonarson, “Ranking Causal Variants and Associated Regions in Genome-Wide Association Studies by the Support Vector Machine and Random Forest,” ***Nucleic Acids Research****,* Vol. 39, No. 9, 2011, pp. e62:1-8, Peer-reviewed. Citations 10.
16. **Zhi Wei\***, W. Wang#, P. Hu, G. J. Lyon, and H. Hakonarson, “SNVer: A Statistical Tool for Variant Calling in Analysis of Pooled or Individual Next-Generation Sequencing Data,” ***Nucleic Acids Research****,* Vol. 39, No. 19, 2011, pp. e132:1-13, Peer-reviewed. Citations 14.
17. W. Wang#, **Zhi Wei**, and Wenguang Sun, “Simultaneous Set-Wise Testing under Dependence, with Applications to Genome-Wide Association Studies,” ***Statistics and Its Interface****,* Vol. 3, No. 4, 2010, pp. 501-512, Peer-reviewed.
18. Hongzhe Li, **Zhi Wei**, and John Maris, “A Hidden Markov Random Field Model for Genome-Wide Association Studies,” ***Biostatistics****,* Vol. 11, No. 1, 2010, pp. 139-150, Peer-reviewed. Citations 8.
19. Caiyan Li, **Zhi Wei**, and Hongzhe Li, “Network-Based Empirical Bayes Methods for Linear Models with Applications to Genomic Data,” ***Journal of Biopharmaceutical Statistics****,* Vol. 20, No. 2, 2010, pp. 209-22, Peer-reviewed. Citations 3.
20. **Zhi Wei**, Kai Wang, Hui-Qi Qu, Haitao Zhang, Jonathan Bradfield, Cecilia Kim, Edward Frackleton, Cuiping Hou, Joseph T Glessner, Rosetta Chiavacci, Charles Stanley, Dimitri Monos, Struan F A Grant, Constantin Polychronakos, and Hakon Hakonarson, “From Disease Association to Risk Assessment: An Optimistic View from Genome-Wide Association Studies on Type 1 Diabetes,” ***PLoS Genetics****,* Vol. 5, No. 10, 2009, pp. e1000678, Peer-reviewed. Citations 71.
21. **Zhi Wei\***, Wenguang Sun, Kai Wang, and Hakon Hakonarson, “Multiple Testing in Genome-Wide Association Studies Via Hidden Markov Models,” ***Bioinformatics****,* Vol. 25, No. 21, 2009, pp. 2802--2808, Peer-reviewed. Citations 13.
22. **Zhi Wei**, Mingyao Li, Timothy Rebbeck, and Hongzhe Li, “U-Statistics-Based Tests for Multiple Genes in Genetic Association Studies,” ***Annals of Human Genetics****,* Vol. 72, No. Pt 6, 2008, pp. 821--833, Peer-reviewed. Citations 13.

**Prior to NJIT**

1. **Zhi Wei**, and Hongzhe Li, “A Hidden Spatial-Temporal Markov Random Field Model for Network-Based Analysis of Time Course Gene Expression Data,” ***Annals of Applied Statistics****,* Vol. 2, No. 1, 2008, pp. 408-429, Peer-reviewed. Citations 30.
2. **Zhi Wei**, and Mingyao Li, “Genome-Wide Linkage and Association Analysis of Rheumatoid Arthritis in a Canadian Population,” ***BMC Proc****,* Vol. 1, No. Suppl 1, 2007, pp. S19, Peer-reviewed. Citations 8.
3. **Zhi Wei**, and Hongzhe Li, “A Markov Random Field Model for Network-Based Analysis of Genomic Data,” ***Bioinformatics****,* Vol. 23, No. 12, 2007, pp. 1537--1544, Peer-reviewed. Citations 92.
4. **Zhi Wei**, and Hongzhe Li, “Nonparametric Pathway-Based Regression Models for Analysis of Genomic Data,” ***Biostatistics****,* Vol. 8, No. 2, 2007, pp. 265--284, Peer-reviewed. Citations 56.
5. **Zhi Wei**, and Shane T. Jensen, “GAME: Detecting Cis-Regulatory Elements Using a Genetic Algorithm,” ***Bioinformatics****,* Vol. 22, No. 13, 2006, pp. 1577--1584, Peer-reviewed. Citations 65.

Citations to my articles (by google scholar)

See <http://scholar.google.com/citations?user=zAva84oAAAAJ&hl=en>



Journal Information for the ones in which I published

The Impact Factor (IF) and ranking information is taken from the 2012 Journal Citation Reports, published annually as part of the Science Citation Index by ISI.

|  |  |  |  |
| --- | --- | --- | --- |
| Journal Name | 2011 IF | 5-yr IF | Note |
| Annals of Applied Statistics | 1.581 | 2.55 | One of the most prestigious Annals serial journals from the Institute of Mathematical Statistics (IMS) |
| Annals of Human Genetics | 2.565 | 2.299 | * Ranks 77th out of 157 journals in *Genetics & Heredity* |
| Bioinformatics | 5.468 | 6.051 | * Ranks 1st out of 47 journals in *Mathematical & Computational Biology* |
| Biostatistics | 2.145 | 3.162 | * Ranks 13th out of 47 journals in *Mathematical & Computational Biology* * Ranks 8th out of 116 journals in *Statistics & Probability* |
| Discovery Medicine | NA | NA |  |
| Frontiers in Biology | NA | NA |  |
| Journal of the American Statistical Association | 1.992 | 3.31 | One of the top four well-recognized statistics journals. The other three are: Annals of Statistics, Biometrika and Journal of the Royal Statistical Society B. |
| Journal of Biopharmaceutical Statistics | 1.342 | 1.338 | * Ranks 30th out of 116 journals in *Statistics & Probability* |
| Journal of Medical Genetics | 6.365 | 5.669 | * Ranks 17th out of 157 journals in *Genetics & Heredity* |
| Nature | 36.28 | 36.235 | Nature |
| Nucleic Acids Research | 8.026 | 7.417 | * Ranks 26th out of 289 journals in *Biochemistry & Molecular Biology* * H5-index ranks 20th among publications across all fields by google scholar. Nature ranks 1st, Science 3rd, JAMA: The Journal of the American Medical Association 12th, Nature Medicine 23rd, IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 97th. See a full top-100 list from: <http://scholar.google.com/citations?view_op=top_venues&hl=en> |
| Pigment Cell Melanoma Research | 5.059 | 5.106 | * Ranks 2nd out of 58 journals in *Dermatology* |
| PLoS Genetics | 8.694 | 9.173 | * Ranks 11th out of 157 journals in *Genetics & Heredity* |
| Physiological Genomics | 2.735 | 3.26 | * Ranks 28nd out of 79 journals in *Physiology* |
| Scientific Reports |  |  | A new primary research publication from the publishers of *Nature,* Not eligible for an impact factor until 2013. |
| Statistics and Its Interface | 0.702 | 0.724 | * Ranks 59th out of 92 journals in *Mathematics, Interdisciplinary Applications* |

**2. Published Refereed Conference Papers**

**At NJIT**

1. Pingzhao Hu, **Zhi Wei**, Zhuozhi Wang, Andrew D. Paterson, Joseph Beyene, and Stephen W Scherer, “Scoring of ChIP-seq experiments by modeling large-scale correlated tests,” ***Critical Assessment of Massive Data Anaysis, CAMDA 2009***, Chicago, Illinois, Vol. 10, No. 1, October 5 - 6, 2009, pp. 25-32, Peer-reviewed.
2. Alexander Braunstein, **Zhi Wei**, Shane T. Jensen, and Jon D. McAuliffe, “A Spatially Varying Two-Sample Recombinant Coalescent, with Applications to Hiv Escape Response,” ***Advances in Neural Information Processing Systems 21, NIPS 2008***, Vancouver, B.C., Canada, Vol. 21, No. 1, Dec. 8 - 13, 2008, pp. 193-200, Peer-reviewed.

**Prior to NJIT**

1. Zhenyu Yan, **Zhi Wei**, and Lishan Kang, ***the International Conference on Artificial Intelligence, IC-AI '03***, Las Vegas, Nevada, Vol. 1, No. 1, June 23 - 26, 2003, pp. 251-256, Peer-reviewed.

**3. Published Refereed Conference Abstracts**

N/A

**C. 1. Published Non-Refereed Journal Papers**

N/A

**2. Published Non-Refereed Conference Papers**

N/A

**3. Published Reports, etc**.

N/A

**4. Published Abstracts** (if not included above)

N/A

**D. Published Reviews**

N/A

**E. Professional Presentations** (Underline Presenter)

**At NJIT**

1. **Zhi Wei**, Bioinformatic analysis for genomic data, HJF Walter Reed Army Institute of Research, Nov. 29, 2012. **Invited**.
2. **Zhi Wei**, W. Wang#, J. Bradfield, E. Frackelton, C. Kim, F. Mentch, R. Baldassano, H. Hakonarson, International IBD Genetics Consortium . Large-sample size, comprehensive catalog of variants and advanced machine learning technique boost risk prediction for inflammatory bowel disease, **Platform talk** (10.6% rate out of 4000+ submissions), The 62th annual meeting of The American Society of Human Genetics (ASHG), San Francisco, CA, Nov 6 - 10, 2012.
3. W. Wang#, **Zhi Wei**. Gene-based Rare-Variant Association Test for Whole Genome Sequencing, the 18th Genetic Analysis Workshop, Stevenson, WA, October 13-17, 2012.
4. C. Kao, **Zhi Wei**, J. Lin, W. Wang#, J. Glessner, C. Cardinale, J. Bradfield, E. Frackelton, C. Kim, F. Mentch, H. Qui, S. Grant, R. Baldassano, H. Hakonarson, International IBD Genetics Consortium, Pathway-Based Meta-Analysis of Ulcerative Colitis Genome-Wide Association Studies, Poster, The 62th annual meeting of The American Society of Human Genetics (ASHG), San Francisco, CA, Nov 6 - 10, 2012.
5. Tao Jiang, Guangqing Sun, Wei Wang#, Jingchu Hu, Paul Bodily, Lifeng Tian, Barry Moore, Hakon Hakonarson, Jun Wang, Mark Yandell, W. Evan Johnson, **Zhi Wei**, Kai Wang, Gholson J. Lyon. Low concordance of variant calling algorithms in exome sequencing, Poster, The 62th annual meeting of The American Society of Human Genetics (ASHG), San Francisco, CA, Nov 6 - 10, 2012.
6. Z. John Daye, Hongzhe Li, **Zhi Wei**, qMSAT: A Powerful Test for Multiple Rare Variants Association Studies That Incorporates Sequencing Qualities, Contributed Paper, Joint Statistical Meetings (JSM), San Diego, CA , July 28 - August 2, 2012.
7. Z. John Daye, Hongzhe Li, **Zhi Wei,** qMSAT: A Powerful Test for Multiple Rare Variants Association Studies that Incorporates, Poster, Sequencing Qualities Conference on New Statistical Methods for Next Generation Sequencing Data Analysis, Ames, Iowa, May 11, 2012.
8. Tao Jiang, Guangqing Sun, Wei Wang#, Jingchu Hu, Paul Bodily, Lifeng Tian, Barry Moore, Hakon Hakonarson, Jun Wang, Mark Yandell, W. Evan Johnson, **Zhi Wei**, Kai Wang, Gholson J. Lyon. Low concordance of variant calling algorithms in exome sequencing, Poster, The Biology of Genomes Conference, Cold Spring Harbor, NY, May 8 - 12, 2012.
9. **Zhi Wei**, Wei Wang, Pingzhao Hu, Lyon GJ, and Hakon Hakonarson, “SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data”, **Platform talk** (21.6% rate out of 1500+ submissions), The 61th annual meeting of The American Society of Human Genetics (ASHG), Montreal, Canada, Oct 11 - 15, 2011
10. **Zhi Wei**, “Statistical methods for analysis of pooled sequencing data”, Joint Statistical Meetings (JSM), **Contributed Topic**, Miami, FL, July 30 - August 4, 2011.
11. G. Zhang, **Zhi Wei**, M. Herlyn, "The requirement of autophagy pathway for drug-induced senescence in human melanoma cells by AZD1152," American Association for Cancer Research 102nd Annual Meeting, Orlando, FL, April 2-6, 2011.
12. **Zhi Wei**, “Statistical methods for analysis of pooled sequencing data”, the SRCOS (Southern Regional Council On Statistics) Meeting, Hickory Knob State Park, South Carolina, June 5-8, 2011. **Invited.**
13. **Zhi Wei**, “Risk Predictions from Genome-Wide Association Data”, the University of Hong Kong, Hong Kong, May 27, 2010, **Invited**.
14. **Zhi Wei**, “An HMM-based Optimal Multiple Testing Procedure for Genome-wide Association Studies”, Biostatistics Branch, National Institute of Environmental Health Sciences, Raleigh, NC, July 31, 2009, **Invited**.
15. **Zhi Wei**, Wenguang Sun, Kai Wang, and Hakon Hakonarson, “A HMM-based optimal Multiple Testing procedure for GWAS”, Poster, The 59th annual meeting of The American Society of Human Genetics (ASHG), Honolulu, Hawaii, Oct 20 - 24, 2009.

**Prior to NJIT**

1. Zhi Wei and Hongzhe Li, Gene Network-based Analysis for Microarray Time Course data in Multiple Biological Conditions**,** Poster,15th Annual International conference on Intelligent Systems for Molecular Biology**,** ISMB2007, Vienna, Austria, July 21-25, 2007
2. Zhi Wei and Hongzhe Li, PMRF: a Pathway-based Markov Random Field Model for Analysis of Genomic Data, Poster, 11th International Conference on Research in Computational Molecular Biology, RECOMB2007, San Francisco, CA, April 21-25, 2007.
3. Hongzhe Li and Zhi Wei, Regularized Estimation in Pathway-Based Censored Data Regression Modeling of Genomic Data, International Biometric Society Conference (ENAR2007), Atlanta, Georgia, March 11-14, 2007.
4. Zhi Wei and Hongzhe Li, Nonparametric Tests of Association of Multiple Genes with Qualitative and Quantitative Phenotypes based on Data-Adaptive U-Statistics, Poster, 5th Asia Pacific Bioinformatics Conference, APBC2007, Hong Kong, Jan. 14-17, 2007.
5. Zhi Wei and Mingyao Li, Genome-wide Association Analysis of Rheumatoid Arthritis in a Canadian Population , 15th Genetic Analysis Workshop (GAW15), St. Pete's Beach, Florida, November 11-15, 2006.

**VI. PROPOSALS AND GRANTS** (Roles must be clearly indicated; include your role and names of others on the proposal/grant: **PI**: Principal Investigator, **CP**: Co-Principal, **I**: Investigator. Summarize in the CV: additional primary information to be submitted in the full dossier including reviews from the granting agencies or grant proposals submitted but not funded.)

#### AWARDED

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Role | PI (if not you) | Agency | Title | Amount | Date of Award |
| NJIT PI | K. Nathanson | Pheopara Alliance | Characterization of somatic genetic and genomic aberrations in inherited pheochro | $8,350 | From  09/01/2011  To  08/31/2012 |
| NJIT PI | M. Herlyn | NIH | Cell-cell communication during melanoma development | $166,442  (out of the total $1,693,507  ) | From  04/01/2011  To  03/31/2016 |
| PI |  | The Children’s Hospital of Philadelphia | Computational analysis of genome-wide association data | $39,720 | From  01/01/2011  To  12/31/2011 |
| PI |  | National Institute of Environmental Health Sciences | Identification and prediction of polyadenylation sites using deep sequencing reads | $6,000 | From  06/01/2010  To  05/31/2011 |

#### PENDING

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Role | PI (if not you) | Agency | Title | Amount | Project Date |
| PI |  | NSF | ABI Innovation: Statistical methods for genomic variant and association analysis | $278,813 | From  05/01/2013  To  04/30/2016 |
| PI |  | NSF | Computational methods for analysis of next-generation sequencing data | $455,665 | From  07/01/2013  To  06/30/2018 |
| PI |  | NSF | BIGDATA: Small: DA: Collaborative Research: Secure, Efficient, and Scalable Methods for Analyzing Big Genetic Association Data | $249,999 | From  02/01/2013  To  01/31/2016 |
| PI |  | NIH | Genetic association methods for next-generation sequencing data | $139,797 | From  04/01/2013  To  03/31/2015 |
| NJIT PI | H. Lee | NIH | Statistical Methods for Identifying the Functional Role of DNA Sequence Variants | $502,583  (out of the total $1,552,392) | From  12/01/2012  To  11/30/2016 |

#### NOT FUNDED

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Role | PI (if not you) | Agency | Title | Amount | Date Submitted |
| PI |  | NIH | Bioinformatics methods for transcription variant study using RNA-Seq | $140,185 | 02/16/2012 |
| PI |  | NIH | NGS Bioinformatics methods for study of ADHD | $373,108 | 06/23/2011  02/27/2012 (revision) |
| PI |  | NSF | SHB:Small:Computational modeling and analysis of high dimensional genomic data | $361,974 | 12/15/2010 |
| PI |  | NIH | Drug-induced senescence in melanoma | $377,006 | 10/22/2010 |
| PI |  | NIH | ABI Innovation: computational modelling and analysis of genetic patterns for studying genetic diseases | $366,032 | 08/19/2010 |
| PI |  | NIH | Disease risk prediction using genome-wide genetic profile | $367,535 | 02/12/2010 |
| Co-PI | Shih, Frank Y. | NSF | Framework of Adaptive Clustering Techniques with Applications to High Dimensional Multimedia and Genetics Data | $776,092 | 02/03/2010 |
| PI |  | NSF | Integrating domain knowledge to build accurate predictive models for analysis of high dimensional genomics data | $338,743 | 12/16/2009 |
| PI |  | NSF | Powerful multiple testing procedures for analysis of correlated high-dimensional data in genome-wide association studies | $253,688 | 08/10/2009 |
| Co-PI | Roshan, Usman W. | NIH | Discriminative SNP selection from genome-wide association studies | $366,972 | 02/22/2009 |
| PI |  | NIH | Optimal Multiple Testing Procedures for Dependent and Grouped Hypotheses in GWAS | $350,429 | 02/17/2009 |

**VII. PATENTS AWARDED (Provide documentation)**

|  |  |  |  |
| --- | --- | --- | --- |
| Author(s) | Title | Patent Number | Date |
| H. Hakonarson, K. Wang, and **Zhi Wei** | Compositions and Methods for Diagnosing Genome Related Diseases and Disorders | 13/499,515 | Submitted on October 8, 2009.  Pending approval. |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**VIII.** **PROFESSIONAL LICENSES**

### Professional Licenses

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Title | State or Country | License Number | Active/Inactive | Date |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

### Certifications

|  |  |  |  |
| --- | --- | --- | --- |
| Title | Certifying Agency | Certification Number | Last Re-certification Date |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**IX. SERVICE ACTIVITIES** (be specific, provide dates; summarize in the CV, additional primary information to be submitted in the full dossier; if the candidate has developed new programs or has played a significant role in developing one, provide data)

1. University

Teaching, Learning and Technology committee, 2011 – 2012

Faculty Marshal for commencement, May 2010

Judge for Graduate Student Research Day, Nov 2009

1. Department

Committee Member

Seminar Committee, Department of Computer Science, 2011 – 2012

Research Committee, Department of Computer Science, 2011 – present

Undergraduate Committee, Department of Computer Science, 2011 – 2012

M. S. Committee, Department of Computer Science, 2008 – present

Department Seminar Organization

11/19//2012 “Machine Learning for Computational Biology”, Yanjun Qi, NEC Labs American

11/12//2012 “Predictive Modeling of Patient State and Therapy Optimization”, Zoran Obradovic, Temple University

09/24//2012 “Microbiome Informatics: Deciphering Microscopic Life and Its Interactions in the Body and the World”, Gail Rosen, Drexel University

04/30//2012 “Computational Problems in Genome Rearrangements”, Kevin Chen, Rutgers University

11/21//2011 “Efficient Full Bayesian HMM for CNV Detection”, Alexander Schliep, Rutgers University

10/31//2011 “The Genetics of Drug-Induced Liver Injury”, Yufeng Shen, Columbia University

09/19//2011 “Genotype Copy Number Variations using Gaussian Mixture Models: Theory and Algorithms”, Qian Kenny Ye, Albert Einstein College of Medicine

04/05/2010 “A Novel Profile-HMM to Predict MicroRNAs and Their Targets Simultaneously”, Junwen Wang, University of Hong Kong

03/08/2010 “Computational Methods for Deciphering Genomic Structures of Bacteria”, Dongsheng Che, East Stroudsburg University of Pennsylvania

04/15/2009 “Hiding Distinguished Ones into Crowd: Privacy-Preserving Publishing Data with Outliers”, Wendy Hui Wang, Stevens Institute of Technology

Open House

Winter 2012, Fall 2011, Winter 2011, Fall 2010, Winter 2009

1. Peer reviewing activity (list papers, proposals, books, etc.)

Journal Papers: (year followed by #articles in brackets if review 1+ articles)

*Bioinformatics (2007, 2008(5), 2009, 2011(2), 2012)*

*BMC Bioinformatics (2011, 2012)*

*BMC Medical genetics (2009)*

*BMC Medical Genomics (2010, 2012)*

*BMC Medical Informatics and Decision Making (2011)*

*BMC Research Notes (2011)*

*BMC Systems Biology (2012)*

*Biostatistics (2010)*

*Human Genetics (2010)*

*IEEE/ACM Transactions on Computational Biology and Bioinformatics (2011)*

*IEEE/ACM Transactions on Evolutionary Computation (2011, 2012(2))*

*International Journal of Computational Biology and Drug Design (2011)*

*International Journal of Data Mining and Bioinformatics (2009)*

*Journal of Computational and Graphical Statistics (2009)*

*Journal of Medical Genetics (2012)*

*Journal of the American Statistical Association (2012)*

*Genome Medicine (2012)*

*Genomics (2012)*

*Journal of Receptor, Ligand and Channel Research*

*Knowledge and Information Systems (2012)*

*Nucleic Acids Research (2012)*

*OMICS: A Journal of Integrative Biology (2009(3))*

*OMICS PUBLISHING GROUP/BIOBIO (2011(2))*

*Statistical Applications in Genetics and Molecular Biology (2011)*

*Statistics in Biosciences (2011)*

*Statistics in Medicine (2008, 2009)*

Proposals

*External proposal review for The Research Grant Council (RGC) of Hong Kong, 2012*

Conference:

*The 4th International Conference on Cooperation and Promotion of Information Resources in Science and Technology (COINFO'09)*

1. Editorial Activity

2012 – present, Editorial Board Member, *Scientifica*

2011 – present, Editorial Review Board Member, *Frontiers in Bioinformatics and Computational Biology*

2011 – present, Editorial Review Board Member, *Frontiers in Applied Genetic Epidemiology*

1. Community/Government
2. Professional Societies
3. Other

**X.** **HONORS, AWARDS, AND LISTINGS** (summarize in the CV, additional primary information to be submitted in the full dossier)

###### XI. STATE YOUR MOST SIGNIFICANT CONTRIBUTIONS TO YOUR FIELD

**SINCE LAST PROMOTION OR TENURE DECISION.**

1. **Pathway-based analysis**. I have developed several innovative methods to utilize biological pathway group-structured information and network-structured information for genomic data analysis. This pioneering work has led to several highly cited publications in the most prestigious journals including *Biostatistics* (Wei and Li 2006, *citations 56*), *Bioinformatics* (Wei and Li 2007, *citations 92*), and *Annals of Applied Statistics* (Wei and Li, 2008 *Citations 30*).
2. **Machine learning methods for risk prediction in genome-wide association studies.** I have been the first one to apply sophisticated machine learning approaches to build accurate disease risk predictive models for genome-wide association studies. Our results for Type 1 Diabetes (T1D) suggest that low-cost SNP genotyping platforms may have the potential to replace the traditional costly HLA-typing methods in assessing T1D risk in clinically relevant settings. We have applied a patent for this work which is pending for approval. These findings were reported by the media Reuters. This work is published in the prestigious genetics journal *PLoS Genetics* in late 2009 and has been cited 70+ times in more than 40 journals including *Nature Review Genetics,* *Nature Reviews Gastroenterology and Hepatology,* and *Nature Genetics.* Our recent results for inflammatory bowel disease will be presented at the coming 62th annual meeting of The American Society of Human Genetics (ASHG) as a platform talk(10.6% rate out of 4000+ submissions).
3. **Large-scale multiple testing**. By modeling dependency, I propose and solve the optimality issue for multiple testing with applications to genomic data (Wei et al *Bioinformatics* 2009, Wang#, Wei, and Sun, *Statistics and Its Interface* 2010, and Sun and Wei, *Journal of the American Statistical Association*, 2011).
4. **Variant call methods for next-generation sequencing data**. I have developed novel statistical methods and software for detecting genomic variants from individual or pooled next-generation sequencing data (Wei et al *Nucleic Acids Research* 2011, *Citations 14*; Wang# et al *Journal of Medical Genetics* 2012). The method was presented in 2011 at the 61th annual meeting of The American Society of Human Genetics (ASHG) as a platform talk(21.6% rate out of 1500+ submissions). Software available at <http://snver.sourceforge.net/> has been downloaded over 750 times so far (November 2012) from more than 35 world-wide countries since it was released in July 2011.

**XII.** **MISCELLANEOUS** (summarize in the CV, additional primary information to be submitted in the full dossier)

1. Annual faculty reports for last five years
2. Any other pertinent activity