## 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: Leung, Ming-Ying

eRA COMMONS USER NAME (credential, e.g., agency login): mleung

POSITION TITLE: Professor of Mathematical Sciences and Director of Bioinformatics and Computational Programs

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 |
|---|---|---|---|
| University of Hong Kong | B.Sc. | 1980 | Mathematics |
| University of Hong Kong | M.Phil. | 1983 | Mathematics |
| Stanford University | M.S. | 1988 | Computer Science |
| Stanford University | Ph.D. | 1989 | Mathematics |

### A. Personal Statement

The major goal of the Border Biomedical Research Center (BBRC) is to facilitate and expand pathobiology research at The University of Texas at El Paso (UTEP) by strengthening the research infrastructure of the institution. The bioinformatics core is responsible for providing bioinformatics computational tools with high-performance computing support to the BBRC investigators, UTEP faculty, and other researchers in the El Paso Region. With broad research interests and experience in probabilistic modeling, algorithm design, and statistical bioinformatics, I have served as PI and co-PI in various projects for setting up and managing computational facilities for biomedical research. Over the past 20 years, I have also been funded by NIH, NSF, IBM, and the Texas Higher Education Coordinating Board to conduct research in viral genome replication and have a long record of publication in this area. In addition, I am involved in several collaborative research projects in identifying protein biomarkers for hepatocellular carcinoma, mammogram image analysis, and HIV vaccine development. As the director of the Bioinformatics and Computational Science Programs, I am familiar with the administrative processes (staffing, budgeting, evaluating, etc.) at UTEP. My expertise in bioinformatics and experience in the management of research facilities, collaborative work, and administration have prepared me to lead the bioinformatics core of BBRC.


1. Leung, M.Y., Burge, C., Blaisdell, B.E., and Karlin, S., (1991) An Efficient Algorithm for Identifying Matches with Errors in Multiple Long Molecular Sequences, *J. Mol. Biol.* 221, 1367-1378. **PMCID: PMC4076298**
2. Leung, M.Y., Marsh, G.M., and Speed, T.P. (1996) Over and Under Representation of Short Oligonucleotides in Herpes Virus Genomes. *J. Computational Biology,* 3(3), 345-360. **PMCID: PMC4076300**
3. Chew, D.S.H., Choi, K.P., Heidner, H., and Leung, M.Y. (2004) Palindromes in SARS and Other Coronaviruses, *INFORMS J. Comp.* 16(4), 331-340. **PMCID: PMC4066412**
4. Schaubert, K.L., Price, D.A., Salkowitz, J.R., Sewell, A.K., Sidney, J, Asher, T.E., Blondelle, S.E., Adams, S., Marincola, F.M., Joseph, A., Sette, A., Douek, D.C., Ayyavoo, V, Storkus, W., Leung, M.Y., Ng, H.L., Yang, O.O., Goldstein, H., Wilson, D.B., Kan-Mitchell, J. (2010) Generation of robust CD8(+) T-cell responses against subdominant epitopes in conserved regions of HIV-1 by repertoire mining with mimotopes. *Eur J Immunol*. 40(7):1950-1962. **PMCID: PMC3086652**


### B. Positions and Honors

**Positions and Employment**

1980-1983: Teaching Assistant, Department of Mathematics, University of Hong Kong
1982-1983: Lecturer, Department of Extramural Studies, University of Hong Kong
1983-1989: Research Assistant and Teaching Fellow, Department of Mathematics, Stanford University

1989-2003 : Assistant and Associate Professor, Division of Mathematics and Statistics, The University of Texas at San Antonio
1993 : Visiting Research Fellow, Department of Statistics, University of California at Berkeley and Department of Pharmaceutical Chemistry, University of California at San Francisco
2001-2002 : Visiting Associate Professor, Department of Statistics, Rice University
2003- : Professor, Department of Mathematical Sciences, and Director of Bioinformatics Program, The University of Texas at El Paso
2013- : Director of Computational Science Program, The University of Texas at El Paso

**Other Experience and Professional Memberships**

1990-1991 : Consultant for the mathematical molecular biology groups at the University of Southern California and Stanford University
2002-2003 : Editorial Board Member, Advances and Applications in Statistics
2002-2004 : Co-chair, Organizing Committee for the International Workshop on Statistical Methods in Microarray Data Analysis, Institute of Mathematical Sciences, National University of Singapore
2005 : Chair, Joint Session in Bioinformatics, 2005 INFORMS Annual Meeting
2005-2009 : External Advisory Board Member, NSF and HHMI funded "Talent Expansion in Quantitative Biology" project at East Tennessee State University
2007-2013 : Associate Editor, INFORMS Journal on Computing
2008 : Chair, Invited paper session in Stochastic Models for Biological Processes, International Workshop on Applied Probability, July 2008, Compiegne, France
2008-2010 : Member, The University of Texas System Computational Biology Workgroup for the Cancer Prevention and Research Institute of Texas
2009- : Member, NIH-RCMI Translational Research Network Translational Informatics Subcommittee
2010 : External Review Panelist for West Virginia IDeA Network for Biomedical Research Excellence, Research Competitiveness Program, American Association for the Advancement of Science
2010 : Chair, Session on Stochastic Models for Biological Systems, 2010 INFORMS Annual Meeting
2011 : Chair for invited Session Computational Methods in Biomolecular and Phylogenetic Analyses, International Federation of Operational Research Societies (IFORS) Conference, July 2011, Melbourne, Australia.
2011 : Chair for invited Cluster on Computational Biology, Institute for Operations Research and Management Science (INFORMS) Annual Conference, November 2011, Charlotte.
2012- : Organizer, Joint UTEP/NMSU Workshop on Mathematics, Computer Science, and Computational Sciences

**Honors**

1986-1987 : Andrew Mellon Foundation Research Award, Institute of Population and Resource Studies, Stanford University
2004 : Professor Y.C. Wong Visiting Lectureship, University of Hong Kong
2007-2008 : Outstanding Performance Award, Office of Research and Sponsored Programs, UTEP

**C. Contribution to Science**

1. Palindromes in Searches for Replication Origins: With my early training and interests in designing efficient algorithms for identifying matches in multiple long molecular sequences, my research focus has been on DNA sequences of Herpesvirus genomes. The most significant contribution is the mathematical characterization of nonrandom clusters of palindromes (e.g., GCAATATTGC), which is a short DNA segment whose reverse complementary sequence is identical to itself. The probability of finding origins of replication around nonrandom clusters of palindromes has been shown to be higher. These replication origins are potential targets for developing vaccines against the growth and spread of viruses. My group continues to find new approaches, such as AT excursion and least-squares support vector machine, to predict the locations of these replication origins more accurately, facilitating the efforts to find targets of vaccine development with less experimentation.
   a. Leung, M.Y., Choi, K.P., Xia, A. and Chen, L.H.Y. (2005) Nonrandom Clusters of Palindromes in Herpesvirus Genomes, *J. Computational Biology* 12(3), 331-354. **PMCID: PMC4032367**
   b. Chew, D.S.H., Choi, K.P., and Leung, M.Y. (2005) Scoring Schemes of Palindrome Clusters for More Sensitive Prediction of Replication Origins in Herpesviruses, *Nucleic Acids Research* 33 (15), e134**. PMCID: PMC1197138**

c. Chew, D.S.H., Leung, M.Y., and Choi, K.P. (2007) AT Excursion: a New Approach to Predict Replication Origins in Viral Genomes by Locating AT-rich Regions. *BMC Bioinformatics* 8, 163-174. **PMCID: PMC1904460**
d. Cruz-Cano, R., Chew, D.S.H., Choi, K.P., and Leung, M.Y. (2010) Least-Squares Support Vector Machine Approach to Viral Replication Origin Prediction, *INFORMS J. Computing*, 22(3), 457-470. **PMCID: PMC2923853**

2. Establishment of RNA Virtual Lab with Databases for RNA Pseudoknots: Palindromes are special cases of the more general patterns of inversions in RNA sequences. Each inversion is a palindrome with a gap between the two complimentary stem sequences. With recent outbreaks of RNA viruses (e.g., SARS, West Niles), our attention has shifted to viruses with RNA molecules as their genomes. Inversions in these RNA molecules have been found to be involved in the formation of stem loops and pseudoknots, sequence patterns important for the formation of their secondary structures and functioning of the viral genomic sequences. Therefore, an RNAVLab has been established for providing a series of software applications and online databases for analyses of RNA secondary structures. As an extension to the existing PseudoBase, we have released the PseudoBase++ version in 2009.
   a. Taufer, M., Leung, M.Y., Solorio, T., Licon, A., Mireles, D., and Johnson, K.L. (2008) RNAVLab: A Virtual Laboratory for Studying RNA Secondary Structures Based on Grid Computing Technology, *Parallel Computing* 34: 661-680. **PMCID: PMC2714649**
   b. Taufer, M., Leung, M.Y., Solorio, T., Licon, A., Mireles, D., and Johnson, K.L. (2008) RNAVLab: Virtual Laboratory for RNA Structure Predictions. Available at http://rnavlab.utep.edu. Accessed 12/19/14.
   c. Taufer, M., Licon, A., Araiza, R., Mireles, D., Gultyaev, A., Van Batenburg, F.H.D., and Leung, M.-Y. (2009) PseudoBase++: An Extension of PseudoBase for Easy Searching, Formatting, and Visualization of Pseudoknots. *Nucleic Acids Research* 37(Database Issue):D127-135. **PMCID: PMC2686561**
   d. Taufer, M., Licon, A., Araiza, R., Mireles, D., Gultyaev, A., Van Batenburg, F.H.D., and Leung, M.-Y. PseudoBase++ for RNA pseudoknots. Available at http://pseudobaseplusplus.utep.edu. Accessed 12/19/14.

3. Long RNA Secondary Structure Prediction Based On RNA Segmentation Algorithm Using Grid Computing: A series of JAVA-based applications and their upgrades have been released in the last few years, e.g., InversFinder 2.0, Segmenta 2.0, and the complete bundle of RNASSA 2.0 has been made available online by 2012 (RNASSA stands for RiboNucleic Acid Secondary Structure Analysis). The core of this software is a new RNA segmentation algorithm based on optimal cuts between inversion clusters along the RNA sequence, relying on the mathematical theory of excursion. With the use of high-throughput grid computing across a network of computers (Bioinformatics Grid) managed by the HTCondor software for task scheduling, we have been able to reduce the computing time from days to a few minutes for structure prediction of RNA over 3000 bases.
   a. Rosskopf, J.J.; Upton, J.H.III, Rodarte, L., Romero, T.A., Leung, M.Y., Taufer, M. and Johnson, K.L. (2010) A 3' terminal stem-loop structure in Nodamura virus RNA2 forms an essential cis-acting signal for RNA replication. *Virus Research* 150(1-2):12-21. **PMCID: PMC3017585**
   b. Mohl, J., Licon, A., Viswakula, S., Kelley, P., Araiza, R., Kodimala, V., Vegesna, R., Saldivar, L., Yehadego, D., Cardenas, G., Vest, E., Taufer, M., Fuentes, O., Johnson, K. L., and Leung, M.-Y. (2012) RNASSA 2.0: RNA Secondary Structure Analysis (Version 2.0.121208). Available at http://rnavlab.utep.edu/rnassa. Accessed 12/19/2014.
   c. Viswakula, S., Kodimala, V., Yehadego, D., Vegesna, R., Taufer, M., Leung, M.-Y., Johnson, K. L., (2012). Segmenta 2.0: A Bioinformatics Tool for RNA Segmentation (Version 2.0.121208). Available at http://rnavlab.utep.edu/rnavlab/downloads.php. Accessed 12/19/2014.
   d. Zhang, B., Yehdego, D. T., Johnson, K. L., Leung, M.-Y., Taufer, M., (2013). Enhancement of accuracy and efficiency for RNA secondary structure prediction by sequence segmentation and MapReduce. *BMC Structural Biology, 13 (Suppl. 1)* (S3), 1-24. Available at www.biomedcentral.com/1472-6807/13/S1/S3. Accessed 12/19/2014. **PMCID: PMC3952952**

4. Extension of the Sequence Segmentation Algorithm for Other Computational-Intensive Problems. With the Translational Bioinformatics Lab and Structural Bioinformatics Lab housing the Bioinformatics Grid with videoconferencing capability established in 2012, we have initiated collaborative projects requiring computationally-intensive tasks and data transfer between remote sites. In collaboration with Dr. Gerken at

Case Western Reserve University on a project entitled “Initiation and Regulation of Mucin Type O-Glycosylation” with specific aims to understand the processes governing O-glycan site selection and O-glycan elongation in order to address the molecular mechanisms and biology of O-glycosylation. We have already constructed a working prototype of a web-based software tool ISOGlyP (Isoform Specific O-Glycosylation Prediction), for predicting O-glycosylation sites in amino acid sequences, available at isoglyp.utep.edu. Using a more sophisticated conceptual model on the Bioinformatics Grid with the high-throughput HTCondor system, our ongoing effort is to extend the current version by incorporating recent data on the N- or C-terminal targeting of previously glycosylated peptide substrate. Other projects include the implementation of a pipeline for predicting GPCRs (G-protein coupled receptors) using hidden Markov models; the application of super-resolution techniques in improving mammograms; and the use of a whole-exome sequencing approach coupled with a bioinformatics gene ontology pipeline to identify exonic DNA variants in patients with ALL (Acute Lymphoblastic Leukemia) from the El Paso Children’s Hospital.

a. Zheng, J., Fuentes, O., Leung, M.-Y., and Jackson, E. (2010) Mammogram Compression Using Super-Resolution. In J. Martí, A. Oliver, J. Freixenet, and R. Martí (Eds.), Vol. 6136, pp. 46-53, Springer-Verlag, Berlin: Lecture Notes in Computer Science. Available at http://www.springerlink.com/content/a63k52581wp84631. Accessed 12/19/14.
b. Cruz-Cano, R., Lee, M.-L. T., Leung, M.-Y., (2012). Logic Minimization and Rule Extraction for Identification of Functional Sites in Molecular Sequences. *BioData Mining, 5*(10), 1-21. Available at www.biodatamining.org/content/5/1/10. Accessed 12/19/2014. **PMCID: PMC3492099**
c. Leung, M.-Y., Cardenas, G. A., Almeida, I. C., Gerken, T. A., (2014). Isoform Specific O-Glycosylation Prediction (ISOGlyP), Version 1.2.. Available at http://isoglyp.utep.edu. Accessed 12/19/2014.
d. Leung, M.-Y., Knapka, J., Oaxaca, D., Harry, W. L., Benjamin, C., Jeremy, R. (2014) Predictive Functional Profiling of Missense Mutations Identified in the Kinome and Phosphatome of High-Risk Leukemia Minority Health and Health Disparities (NIMHD) Grantees' Conference, National Harbor, MD.

**Complete List of Published Work in MyBibliography: http://www.ncbi.nlm.nih.gov/sites/myncbi/ming-ying.leung.1/bibliograpahy/47215158/public/?sort=date&direction=ascending**

**D. Research Support**

**Ongoing Research Support**

Grant: 2G12MD007592 (Kirken) 7/1/2014 – 3/31/2019
Source: NIH
Title: Border Biomedical Research Center
Major Goal: Achieving the RCMI-BBRC goal and objectives will help create a healthier Paso de Norte border region by spearheading, sustaining and leveraging biomedical research and practice, and increasing the work force to develop countermeasures to health problems concentrated in the area but of national consequence, and provide a highly competitive faculty and training setting at UTEP to prepare minority students for entry into the biomedical research mainstream of the nation.
Role: Director, Bioinformatics Core Facility

Grant: 1R15AI105823-01A1 (Johnson) 4/1/2014 – 3/31/2017
Source: NIH
Title: Mechanisms in Viral RNA Replication Complex Assembly: Novel Targets For Antivira
Major Goal: This study will uncover essential processes in viral RNA replication common to many pathogenic viruses and will identify new targets for antiviral therapies.
Role: Co-I

Grant: 2012-38422-19910 (Leung) 9/1/12 – 8/31/15
Source: USDA
Title: Bioinformatics Education for Agricultural Science
Major Goal: The major goals of this project are to support undergraduates and graduate students, and train them to conduct bioinformatics analysis of data from USDA research. The project will generate educational materials packaged in the form of web-based modules that can be readily utilized for lecture presentation and class projects in bioinformatics-related courses in biology, biochemistry, computer science, and mathematics.
Role: PI

Grant: DUE 0966151 (Pannell) 9/1/2010 – 8/31/2015
Source: NSF
Title Recruiting and Keeping Undergraduate Students in the Sciences
Major Goal: The major goal of this project is to offer scholarships to talented freshmen majoring in biology, chemistry, mathematics, geology, or physics.
Role: Co-PI

Grant: DUE 0926721 (Leung) 9/1/2009 – 8/31/2015
Source: NSF Interdisciplinary Training for Undergraduates in Biological and Mathematical Sciences (UBM)
Title: UBM Institutional: Undergraduate Training in Bioinformatics
Major Goal: The major goal of this project is to establish an undergraduate training program of bioinformatics research at UTEP.
Role: PI

**Completed Research Support**

Grant: 2G12RR008124 (Natalicio) 9/7/2009 – 6/30/2014
Source: NIH/NCRR
Title: Border Biomedical Research Center
Major Goal: The major goal of this project is to facilitate and expand the pathobiology research at The University of Texas at El Paso (UTEP) by strengthening the research infrastructure of the institution.
Role: Director, Bioinformatics Computing Laboratory (BCL) Core Facility

Grant: 1T36GM078000-01 (Aley) 6/1/2008 – 5/31/2014
Source: NIH Minority Access to Research Careers (MARC) Program
Title: Enhancement of Quantitative Science in Biology Curricula
Major Goal: The major goals of this project are to develop a revised biology curriculum with a strong, quantitative framework and with integral interdisciplinary ties to mathematical science and computer science.
Role: Co-PI

Grant: DMS 0800272 (Leung) 6/1/2008 – 5/31/2013
Source: NSF
Title: Mathematical Models for RNA
Major Goal: The major goals of this project are to develop mathematical models for accurate and consistent prediction of RNA secondary structures in large RNA molecules and to test the validity of the predicted structures by mutagenesis experiments using the genomes of nodaviruses as models.
Role: PI

Grant: 1R01AI077413 (Kan-Mitchell) 9/1/2008 – 8/31/2012
Source: NIH
Title: Mapping Novel Subdominal B*5701 Epitopes in Conserved Regions of the HIV Proteome
Major Goal: The major goal of this project is to map subdominant HIV epitopes in candidate regions of the HIV proteome under great biological constraints that are presented by B*5701, an allele closely associated with HIV control.
Role: Co-I