

SOHRAB P SHAH CURRICULUM VITAE

Contact Information

Centre for Translational and Applied Genomics & Molecular Oncology and Breast Cancer Research Program, BC Cancer Agency

Email: sshah@bccrc.ca

Phone: +1 604-877-6000 x2589

Mobile: +1 778-885-1365

Fax: +1 604-877-6089

Mail: rm 3427 600 W10th Ave. Vancouver BC V5Z4E6 CANADA

Web: http://compbio.bccrc.ca/?page_id=93

Education

PhD in Computer Science (bioinformatics), University of British Columbia (2008)

Supervisors: Drs. Raymond Ng and Kevin Murphy

MSc in Computer Science (bioinformatics), University of British Columbia (2005)

Supervisor: Dr. Anne Condon

BSc in Computer Science, University of British Columbia (2001)

BSc(Hons) in Biology, Queen's University (1996)

Work Experience

Postdoctoral Research Fellow, BC Cancer Agency (May 2008-present)

Research Assistant, UBC Dept of Computer Science (May2005-May2008)

Instructor, Interprofessional Health and Human Services, UBC (July 2006-Aug2009)

Instructor, Canadian Bioinformatics Workshops Series (2002-present)

Chief, High throughput bioinformatics, UBC (May2002-Aug2004)

Bioinformatics software developer, Centre for Molecular Medicine and Therapeutics (May2000-May2002)

Awards and distinctions

Lap-Chee Tsui Publication Award from the Canadian Institutes for Health Research Institute of Genetics in recognition of outstanding published health research carried out by trainees, For the discovery of the mutation in *FOXL2* in granulosa cell tumors of the ovary (published in *NEJM*). (\$1000).

Research Fellowship, Eli Lilly. (\$130,000 over 2yrs) Effective 09/2009

Postdoctoral Fellowship, Michael Smith Foundation For Health Research, Canada (\$120,000 over 3yrs) Effective 10/2008-10/2011

Canadian Breast Cancer Foundation Bioinformatics Fellowship. (part of \$500,000 over 5yrs to Dr. Sam Aparicio)

ISCB travel award, International Society for Computational Biology, United States, Effective: 07/2009, Ending: 08/2009, Travel fellowship to ISMB 2009 conference, \$1,100

SOHRAB P SHAH CURRICULUM VITAE

ISCB travel award, International Society for Computational Biology, United States, Effective: 07/2007, Ending: 08/2007, Travel fellowship to ISMB 2007 conference, \$1,100

University Graduate Fellowship (declined), University of British Columbia, Canada, Effective: 05/2006, Ending: 05/2009, \$48,000

Senior Graduate Trainee Award, Michael Smith Foundation For Health Research, Canada, Effective: 04/2006, Ending: 04/2008, \$45,000

Student service award. Department of Computer Science, UBC 2008.

Funding

Canadian Institute for Health Research: Bioinformatics catalyst grant. Predicting single nucleotide variants from next generation sequencing of cancer genomes. \$100,000/yr for 1yr. Nov 2009-Nov 2010. Co-investigator and author of the grant. PI: Sam Aparicio. **Ranked 1st out of 17 applications.*

Publications

Published refereed papers (selected)

1. Goya R, Sun M, Morin RD, Leung G, Ha G, Wiegand KC, Senz J, Crisan A, Marra MA, Hirst M, Huntsman DG, Murphy KP, Aparicio S, **Shah SP**. SNVMix: predicting single nucleotide variants from next generation sequencing of tumors. *Bioinformatics*. In Press. *Contribution: Project conception, study design and oversight, corresponding author.*
Citations: (manuscript in press)
2. **Shah SP**, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, vol**461**, 809-813. (2009).
Contribution: project lead. Developed algorithms, performed analysis and identified mutation candidates for experimental validation. Co-wrote the manuscript
Citations: 2 (Google Scholar)
Note: this article was featured on the cover of this issue of Nature*
3. **Shah SP**, Köbel M, Senz J, Morin RD, Clarke BA, Wiegand KC, Leung G, Zayed A, Mehl E, Kalloger SE, Sun M, Giuliany R, Yorida E, Jones S, Varhol R, Swenerton KD, Miller D, Clement PB, Crane C, Madore J, Provencher D, Leung P, DeFazio A, Khattra J, Turashvili G, Zhao Y, Zeng T, Glover JN, Vanderhyden B, Zhao C, Parkinson CA, Jimenez-Linan M, Bowtell DD, Mes-Masson AM, Brenton JD, Aparicio SA, Boyd N, Hirst M, Gilks CB, Marra M, Huntsman DG. Mutation of FOXL2 in granulosa-cell tumors of the ovary. *N Engl J Med*. 2009 Jun 25;**360(26)**:2719-29.

SOHRAB P SHAH CURRICULUM VITAE

Contribution: data analysis and algorithm development. Interpretation of results that led to the mutation discovery. Co-wrote the manuscript

Citations: 13 (Google Scholar) including 5 review articles on next generation sequencing

Note: winner of Lap-Chee Tsui Publication Award from the Canadian Institutes for Health Research Institute of Genetics*

4. Steidl C, Lee T, **Shah SP**, Farinha P, Han G, Nayar T, Delaney A, Jones SJ, Bast MA, Iqbal J, Rosenwald A, Rimsza LR, Campo E, Jaffe ES, Lenz G, Connors JM, Staudt LM, Chan WC, Gascoyne RD. Tumor-associated macrophages predict long-term survival in classical Hodgkin lymphoma. *N Engl J Med*. In Press.
Contribution: supervised data analysis and bioinformatics component of the study
Citations: (manuscript in press)
5. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F, Yap D, Humphries RK, Griffith OL, **Shah SP**, et al. Somatic mutation of EZH2 (Y641) in Follicular and Diffuse Large B-cell Lymphomas of Germinal Center Origin. *Nature Genetics*. In press
Contribution: minor role in data analysis
Citations: (manuscript in press)
6. **Shah SP**, Cheung KJ Jr, Johnson NA, Alain G, Gascoyne RD, Horsman DE, Ng RT, Murphy KP. Model-based clustering of array CGH data. *Bioinformatics*. 2009 Jun 15;25(12):i30-8.
Contribution: derived, implemented and evaluated the algorithm. Wrote the manuscript.
Citations: 2 (Google scholar)
7. **Shah SP**. Computational methods for identification of recurrent copy number alteration patterns by array CGH. *Cytogenet Genome Res*. 2008;123(1-4):343-51.
Contribution: sole author solicited review article
Citations: 2 (Google scholar)
8. Cheung KJ*, **Shah SP***, Steidl C, Johnson N, Relander T, Telenius A, Lai B, Murphy KP, Lam W, Al-Tourah AJ, Connors JM, Ng RT, Gascoyne RD, Horsman DE. Genome-wide profiling of follicular lymphoma by array comparative genomic hybridization reveals prognostically significant DNA copy number imbalances. *Blood*. 2008 Aug 14.
(*equal contribution)
Contribution: led and performed data analysis
Citations: 8 (Google scholar)
9. **Shah SP**, Lam WL, Ng RT, Murphy KP. Modeling recurrent DNA copy number alterations in array CGH data. *Bioinformatics*. 2007 Jul 1;23(13):i450-8.
Contribution: derived, implemented and evaluated the algorithm. Wrote the manuscript.
Citations: 28 (Google scholar)

SOHRAB P SHAH CURRICULUM VITAE

10. **Shah SP**, Xuan X, DeLeeuw RJ, Khojasteh M, Lam WL, Ng R, Murphy KP. (2006) Integrating copy number polymorphisms into array CGH analysis using a robust HMM. *Bioinformatics*. Jul 15;22(14):e431-9.
Contribution: derived, implemented and evaluated the algorithm. Wrote the manuscript.
Citations: 39 (Google scholar)
11. Kemmer D, Huang Y, **Shah SP**, Lim J, Brumm J, Yuen MSM, Ling J, Xu T, Wasserman WW, Ouellette BF. (2005) Ulysses - an Application for the Projection of Molecular Interactions across Species. *Genome Biology*. 6(12).
Contribution: developed the Atlas data warehouse enabled this analysis
Citations: 19 (Google scholar)
12. **Shah SP**, Huang Y, Xu T, Yuen MM, Ling J, Ouellette BF. (2005) Atlas - a data warehouse for integrative bioinformatics. *BMC Bioinformatics* 21;6(1):34
Contribution: designed, implemented and tested the software. Wrote the manuscript.
Citations: 71 (Google scholar)
13. **Shah SP**, He DY, Sawkins JN, Druce JC, Quon G, Lett D, Zheng GX, Xu T, Ouellette BF. (2004) Pegasys: software for executing and integrating analyses of biological sequences. *BMC Bioinformatics* 5(1):40.
Contribution: designed, implemented and tested the software. Wrote the manuscript.
Citations: 56 (Google scholar)
14. **Shah SP**, McVicker GP, Mackworth AK, Rogic S, Ouellette BF. (2003) GeneComber: combining outputs of gene prediction programs for improved results. *Bioinformatics* 19(10):1296-7.
Contribution: designed, implemented and tested the software. Wrote the manuscript.
Citations: 17 (Google scholar)

Submitted papers

Christian Steidl*, **Sohrab P. Shah***, Bruce Woolcock, Pedro Farinha, Nathalie A. Johnson, Yongjun Zhao, Adele Telenius, Susana Ben Neriah, Arjan Diepstra, Anke van den Berg, Mark Sun, Gillian Leung, Joseph M. Connors, David G. Huntsman, Kerry J. Savage, Lisa Rimsza, Douglas E. Horsman, Marco A. Marra and Randy D. Gascoyne. Recurrent rearrangement of MHC class II transactivator CIITA in B cell lymphomas. Submitted. Role: Led data analysis (*-equal contribution).

Published contributions to a collective work

Khojasteh M, Coe B, **Shah SP**, Ward RK, Lam WL, MacAulay C. A Novel Algorithm for the Analysis of Array CGH Data. Submitted to 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (Confirmation of submission attached).

role: collaborator, 5% contribution

SOHRAB P SHAH CURRICULUM VITAE

Sohrab P Shah, Jessica N Sawkins, Macaire MS Yuen, Joanne A Fox, David Arenillas, Pamela J Thomas, John Ling, NISC Comparative Sequencing Program, Eric D Green, Wyeth W Wasserman, BF Francis Ouellette. High-Quality Annotation from Twelve Vertebrates for Genes Orthologous to a Focused Region of Study in the Human Genome. In Proceedings of *Genome Informatics 2005*, Cold Spring Harbour Laboratories, USA.

Role: Project lead, 75% contribution

Shah SP, He DY, Sawkins JN, Ouellette BF. (2003) Pegasys: a Parallel Genome Annotation System. In Proceedings of *Genome Informatics 2003*, Cold Spring Harbour Laboratories, USA. (platform presentation - presented by **SP Shah**)

Role: Project lead, 75% contribution

Shah SP, Venkatesh B, Major F, Ouellette BF. (2001) Finding Conserved Patterns in Fugu Introns. In Proceedings of *Mathematical Formalisms in RNA Sequences 2001*, Montreal, Canada.

Role: Project lead, 95% contribution

Theses

Shah SP. Model based approaches to array CGH data analysis. PhD dissertation. Dept of Computer Science, UBC. Nov 2008

Shah SP. Detecting common secondary structure elements in RNA sequences. MSc Thesis. Dept of Computer Science, UBC. May 2005

Activities and Contributions

Committee Membership

Associate Editor. *BMC Cancer* as of Jan 2010

Scientific officer and reviewer. Canadian Cancer Society Research Institute. 2009 Panel J2 competition.

Scientific committee for the NETTAB2005 workshop entitled: "Workflows management: new abilities for the biological information overflow", Oct5-7, 2005. Naples, Italy

Subject: workflows in bioinformatics

Evaluation of Articles

Referee for *Biometrika*, Wellcome Trust, Bioinformatics, NAR, Pacific Symposium on Biocomputing, *Journal of Statistical Methodology*, *BMC Bioinformatics*, *BMC Genomics*, *PLoS Computational Biology*

Supervisory experience

Currently supervising 3 M.Sc. students doing rotation projects as part of the CIHR Bioinformatics training program

SOHRAB P SHAH CURRICULUM VITAE

Presentations as guest speaker/Participation in training workshops

Invited Speaker: "Mutation discovery in cancer using next generation sequencing and probabilistic models". Vancouver Bioinformatics Users Group Seminar Series, Vancouver, BC. Nov 26, 2009.

Invited Speaker: "Mutation discovery in cancer using next generation sequencing and probabilistic models". Stanford University. Nov 13, 2009. Host: Matt van de Rijn

Invited Speaker: "Mutation discovery in cancer using next generation sequencing and probabilistic models". Centre for Cellular and Biomolecular Research, University of Toronto. Nov 3, 2009. Host: Michael Brudno

Conference oral presentation: "Probabilistic models for detecting SNVs from next gen sequencing of tumours". ISMB 2009 Stockholm, Sweden

Conference oral presentation: "Model-based clustering of array CGH data". ISMB 2009 Stockholm, Sweden.

Conference oral presentation: "Computational inference of copy number changes from next generation sequencing of cancer genomes." AGBT 2009, Feb 6, 2009 Marco Island, USA

Conference oral presentation: "Modeling recurrent DNA copy number alterations in array CGH data." ISMB 2007, July 23, 2007 Vienna, Austria

Conference oral presentation: "Integrating copy number polymorphisms into array CGH analysis using a robust HMM." ISMB 2006, July 20, 2006 Fortaleza, Brazil

Invited Speaker. NETTAB2005: "Workflows management: new abilities for the biological information overflow", Naples, Italy. Oct 7, 2005 Pegasys: workflow management for genome annotation.

Education/Teaching

Guest Lecturer: UBC Statistics STAT547M Topics in Statistics. Jan 14, 2009.

Lead faculty: "Clinical genomics and biomarker discovery". Canadian Bioinformatics Workshops Series. July 16-17, 2009 Toronto, ON. Developed and delivered curriculum.

Guest lecturer: IHHS 302: Topics in Health Informatics for Health/Life Science Students. Topic: Genomics and bioinformatics. Aug, 2009.

Instructor: IHHS 302: Topics in Health Informatics for Health/Life Science Students. 3rd yr level UBC course. Aug 2005, 2006, 2007

SOHRAB P SHAH CURRICULUM VITAE

Instructor, Canadian Bioinformatics Workshops Series. Genomics. July, 2007, Vancouver BC.
Analysis of array CGH data

Instructor. Canadian Bioinformatics Workshops Series. Vancouver, Canada. Feb 18, 2005.
Introduction to Programming for Bioinformatics.

Instructor. Canadian Bioinformatics Workshops Series. Montreal, Canada. 2005.
Bioinformatics: Developing the Tools

Instructor, UBC Library Science LIBR548D. Nov 28, 2003. What is bioinformatics?