



FORM 100
Personal Data Form
PART I

Date
2008/10/20

Family name Ng	Given name Raymond	Initial(s) of all given names T.	Personal identification no. (PIN) 109895
-------------------	-----------------------	-------------------------------------	---

I hold a faculty position at an eligible Canadian college (complete Appendices B1 and C)

I do not or will not hold an academic appointment at a Canadian postsecondary institution

Place of employment other than a Canadian postsecondary Institution (give address in Appendix A)

APPOINTMENT AT A POSTSECONDARY INSTITUTION

Title of position Professor	Tenured or tenure-track academic appointment Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
Department Computer Science	Part-time appointment <input type="checkbox"/> Full-time appointment <input checked="" type="checkbox"/>
Campus Vancouver, BC V6T 1Z4	<ul style="list-style-type: none"> For all non-tenured or non tenure-track academic appointment and Emeritus Professors, complete Appendices B & C For life-time Emeritus Professor and part-time positions, complete Appendix C
Canadian postsecondary institution British Columbia	

ACADEMIC BACKGROUND

Degree	Name of discipline	Institution	Country	Date yyyy/mm
Bachelor's	Computer Science	University of British Columbia	CANADA	1986/05
Master's	Computer Science	University of Waterloo	CANADA	1988/05
Doctorate	Computer Science	University of Maryland	U.S.A.	1992/05

TRAINING OF HIGHLY QUALIFIED PERSONNEL

Indicate the number of students, fellows and other research personnel that you:

	Currently		Over the past six years (excluding the current year)		Total
	Supervised	Co-supervised	Supervised	Co-supervised	
Undergraduate					
Master's	1	2	15	2	20
Doctoral	3		2	1	6
Postdoctoral	1	1	1	3	6
Others					
Total	5	3	18	6	32

Personal identification no. (PIN)

109895

Family name

Ng

ACADEMIC, RESEARCH AND INDUSTRIAL EXPERIENCE (use one additional page if necessary)

Position held (begin with current)	Organization	Department	Period (yyyy/mm to yyyy/mm)
Professor	British Columbia	Computer Science	2002/07
Associate Professor	University of British Columbia	Computer Science	1997/07 to 2002/06
Assistant Professor	University of British Columbia	Computer Science	1992/07 to 1997/06
Research Associate	Xerox Webster Research Center		1991/05 to 1991/08
Research Associate	University of Waterloo		1988/04 to 1988/07
Summer Programmer	University of British Columbia		1985/05 to 1985/08

Personal identification no. (PIN)

109895

Family name

Ng

RESEARCH SUPPORT

Family name and initial(s) of applicant	Title of proposal, funding source and program, and time commitment (hours/month)	Amount per year	Years of tenure (yyyy)
List all sources of support (including NSERC grants and university start-up funds) held as an applicant or a co-applicant: a) support held in the past four (4) years but now completed; b) support currently held, and c) support applied for. For group grants, indicate the percentage of the funding directly applicable to your research. Use additional pages as required.			
a) Support held in the past 4 years			
Ng, R., Little, J. et al.	Acquisition, Querying and Prediction of Motion Trajectories NCE IRIS 15 hours/month	190,000 (15%) 190,000 (15%) 190,000 (15%)	2002 2003 2004
Lakshmanan, L., Ng, R. et al.	New Frontiers in Data Mining NCE IRIS 10 hours/month	145,000 (20%) 145,000 (20%) 145,000 (20%)	2002 2003 2004
Jurisica, I., Glasgow, J., Hoos, H. and Ng. R.	Intelligent Computational Methods for the Analysis of Gene Expression Profiles NCE IRIS 15 hours/month	193,000 (20%) 193,000 (20%) 193,000 (20%)	2002 2003 2004
Raymond Ng	Robust tools for biomedical data NSERC Equipment grant 4 hours/month	48,686(100%)	2004

Personal identification no. (PIN)

109895

Family name

Ng

RESEARCH SUPPORT

Family name and initial(s) of applicant	Title of proposal, funding source and program, and time commitment (hours/month)	Amount per year	Years of tenure (yyyy)
List all sources of support (including NSERC grants and university start-up funds) held as an applicant or a co-applicant: a) support held in the past four (4) years but now completed; b) support currently held, and c) support applied for. For group grants, indicate the percentage of the funding directly applicable to your research. Use additional pages as required.			
a) Support held in the past 4 years			
Bruce McManus, Paul Keown, Rob McMaster, R Ng et a	Better Biomarkers of Acute and Chronic Allograft Rejection Genome Canada Applied Health 8 hours/month	3,033,000 (1%) 3,033,000 (1%) 3,033,000 (1%)	2005 2006 2007
Stephen Lam, Victor Ling, Calum McCaulay, R. Ng et	Pharmacogenomics for Rational Chemotherapy of Lung Cancer Genome Canada Applied Health 4 hours/month	2,300,000 (1%) 2,300,000 (1%) 2,300,000 (1%)	2005 2006 2007
Doug Horsman, Wan Lam, Calum McCaulay, R Ng et al.	Development and validation of Comparative Genomic Hybridization Arrays for Clinical Use in Cancer Genome Canada Applied Health 2 hours/month	1,500,000 (0%) 1,500,000 (0%) 1,500,000 (0%)	2005 2006 2007
Giuseppe Carenini and Raymond Ng	Summarizing of evaluative text and emails Google University scholarship 4 hours/month	34,000 (50%)	2007

Personal identification no. (PIN)

109895

Family name

Ng

RESEARCH SUPPORT

Family name and initial(s) of applicant	Title of proposal, funding source and program, and time commitment (hours/month)	Amount per year	Years of tenure (yyyy)
List all sources of support (including NSERC grants and university start-up funds) held as an applicant or a co-applicant: a) support held in the past four (4) years but now completed; b) support currently held, and c) support applied for. For group grants, indicate the percentage of the funding directly applicable to your research. Use additional pages as required.			
b) Support currently held			
Raymond Ng	Robust Mining of Biomedical Data NSERC discovery grant, individual 16 hours/month	54,700(100%) 54,700(100%) 54,700(100%) 54,700(100%) 54,700(100%)	2004 2005 2006 2007 2008
C. Helbing, C. Borchers, M. Lesperance, R. Ng	New generation tools for the assessment of nanomaterial effects on amphibian wildlife NSERC strategic projects 12 hours/month	249,880 (15%) 232,880 (15%) 233,880 (15%)	2009 2010 2011
Renee Miller, Raymond Ng and 12 others	Business Intelligence Network NSERC strategic network 16 hours/month	1,000,000 (5%) 1,000,000 (5%) 1,000,000 (5%) 1,000,000 (5%) 1,000,000 (5%)	2009 2010 2011 2012 2013

Highly Qualified Personnel (HQP)

Provide personal data about the HQP that you currently, or over the past six years, have supervised or co-supervised.

			Personal identification no. (PIN)	Family name
			109895	Ng
Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
Gabriel Murray	Postdoctoral (In Progress)	Co-supervised 2008 - 2010	multi-modal text summarization	postdoctoral fellow
Kristic McBrunie	Master's (In Progress)	Supervised 2007 - 2009	Analysis of Gene Expression of Contaminants	graduate student
Hongae Lee	Doctoral (In Progress)	Supervised 2006 - 2009	selectivity estimation with edit distance	graduate student
Gariella Cohen-Freue	Postdoctoral (In Progress)	Co-supervised 2005 - 2009	Analysis of Proteomics Expression	research associate
Jan Ulrich	Master's (Completed)	Co-supervised 2006 - 2008	Extractive summarization of emails	founding a startup company
Shaofeng Bu	Doctoral (In Progress)	Supervised 2005 - 2008	Privacy preserving data mining	graduate student
Sohrab Shah	Doctoral (In Progress)	Co-supervised 2005 - 2008	Breakpoint detection for CGH arrays	graduate student
Xiaodong Zhou	Doctoral (Completed)	Co-supervised 2003 - 2007	Summarization of Email Folders	research staff member, AOL Research
Ming Yue	Master's (Completed)	Supervised 2004 - 2006	Comparative study of kernel based classification and feature	research staff member, Microsoft research
Ganesh Ramesh	Postdoctoral (Completed)	Co-supervised 2003 - 2005	Foundations of XML Querying	data mining analyst, Microsoft research
Shaofeng Bu	Master's (Completed)	Supervised 2003 - 2005	GMDL Generalization with Exceptions	graduate student
Timothy Chan	Master's (Completed)	Supervised 2003 - 2005	Permutation tests for Early Lung Cancer Detection	data mining analyst
Zuzanna Hollander	Master's (Completed)	Supervised 2003 - 2005	Gene expression profiling for rheumatic valves	data mining analyst
Rong Zhu	Postdoctoral (Completed)	Co-supervised 2002 - 2004	Covariance Analysis of Genes	postdoctoral fellow
(Name withheld)	Master's (Completed)	Co-supervised 2002 - 2003	Mining of Bug Reports for Software Maintenance	research staff
(Name withheld)	Master's (Completed)	Co-supervised 2002 - 2003	Detection of Activation and Inhibiting Pairs of Genes	research analyst
Ruiyao Yang	Master's (Completed)	Supervised 2002 - 2003	Summarization of Time Series	software developer, Crystal Decisions
Carson Leung	Doctoral (Completed)	Supervised 2000 - 2003	Constrained Mining of Frequent Sets	Assistant Professor, Computer Science, U. Manitoba
Fatemah Alqallaf	Doctoral (Completed)	Co-supervised 1999 - 2003	A new Contamination Model for Robust Estimation	Assistant Professor, Statistics, Kuwait University
Joerg Sander	Postdoctoral (Completed)	Supervised 2001 - 2002	Clustering of Genes based on Expression	Assistant Professor, Computer Science, U. Alberta

Highly Qualified Personnel (HQP)

Provide personal data about the HQP that you currently, or over the past six years, have supervised or co-supervised.

			Personal identification no. (PIN) 109895	Family name Ng
Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
Min Phan	Master's (Completed)	Supervised 2001 - 2002	A Tool for Managing Genomic Data	software developer, Crystal Decisions
Steven Wang	Postdoctoral (Completed)	Co-supervised 2001 - 2002	Robust Methods for Data Mining	Assistant Professor, Statistics, U. Ottawa
Xiaodong Zhou	Master's (Completed)	Supervised 2001 - 2002	GMDL Generalization of OLAP Cells	research staff member, AOL Research
Ed Knorr	Doctoral (Completed)	Supervised 1999 - 2002	Detection of distance-based outliers	Instructor, Computer Science, UBC
(Name withheld)	Master's (Completed)	Supervised 2000 - 2001	mining of sequential patterns	software developer

Form 100 (2008 W), page 4-1 of 4

Personal information collected on this form and appendices will be stored in the Personal Information Bank for the appropriate program.

Version française disponible

Canada

PROTECTED WHEN COMPLETED

Research Contributions

1. Most Significant Contributions to Research

In [24], we developed a new clustering algorithm designed for large databases. This algorithm, when first published in the proceedings of the 1994 VLDB conference, was the first one of its kind, and inspired many recent attempts in the same direction. Consequently, in the past few years, the conference version is cited in almost every data mining paper where clustering is a relevant topic. Furthermore, the software we developed has become a standard benchmark. Fellow researchers developing new clustering algorithms for data mining applications often compare their algorithms with ours. According to Google Scholar, there are over 1200 citations of this paper so far.

Genomic data analysis is very sensitive to data pre-processing. For example, breakpoint detection is crucial to the analysis of CGH arrays. In [9] and [17], we developed a robust framework based on Hidden Markov Models to model DNA copy number alterations. Both papers were accepted to the ISMB conference, with the full version appearing in the Bioinformatics journal. The Bioinformatics journal, with an impact factor of 6, is the highest ranked journal for bioinformatics research. Moreover, for microarrays, various quality control checks are indispensable for gene expression analysis. In [8] and [15], we proposed two different methods that provide enhanced quality control of microarray data.

As for actual genomic analysis, we have published many papers on various types of cancer. The most talked-about paper is [7], which focuses on smoking and lung cancer. One of the key findings is that even though a smoker may quit smoking for years or even decades, certain gene-repairing mechanisms are permanently damaged. Immediately after the paper was published, some of us were on Canada wide TV and interviewed by North American newspapers including the Los Angeles Times and Chicago Tribune.

In [42], we developed an efficient indexing method for multi-dimensional trajectories. The paper won the 2004 ACM SIGMOD best paper award. ACM SIGMOD is one of the top two database conferences in the world. The acceptance rate was 69 out of 431. Moreover, the paper entitled “Robust Space Transformations for Distance-based Operations” won the 2001 ACM SIGKDD best paper award. ACM SIGKDD is the most prestigious data mining conference in the world. The acceptance rate for 2001 SIGKDD was 20 out of 203. The paper makes contributions to both computational statistics and databases. Specifically, the techniques we proposed make many database and data mining operations more meaningful and robust against outlying values.

In [29], [30] and [31], we proposed a comprehensive framework for evaluative text. We first considered the problem of text summarization based on a hierarchical value system. We also considered multi-document summarization where improved fluency is one of the main issues addressed. Finally, we designed and developed an interactive multimedia environment for a reader to read the summaries, as well as to explore the supporting text. This line of work can find many business applications such as summarizing customer reviews and user complaints. Another component of our text mining research concerns email summarization. In [25], [33] and [37], we studied the problem of summarizing large folders of emails. These two text mining projects are partially funded by Google.

2. List of Research Contributions in the Last 6 Years

(For Computer Science papers, the default ordering of authors is alphabetical. When it is not, the authors are listed in decreasing order of perceived contribution. For medical related papers, the last few authors are the most senior authors. Students or post-doctoral fellows are bold-faced.)

A. Refereed Journal Publications (2002-2008)

- [1] **Chari, R.**, Coe, B., Wedseltoft, C., Benetti, M., Wilson, I., Vucic, E., MacAulay, C., Ng, R. and Lam, W. "SIGMA2: A system for the integrative genomic multi-dimensional analysis of cancer genomes, epigenomes, and transcriptomes," To appear in: *BMC Bioinformatics*.
- [2] Tsui, I., Rosin, M., Zhang, L., Ng, R. and Lam, W. "Multiple aberrations of chromosome 3p detected in oral premalignant lesions," To appear in: *Cancer Prevention Research*.
- [3] Lakshmanan, L., Ng, R. and **Ramesh, G.** "On Disclosure Risk Analysis of Anonymized Itemsets in the Presence of Prior Knowledge," To appear in: *ACM Transactions on Knowledge Discovery from Data*.
- [4] Cheung, K.J., **Shah, S.**, Ng, R. et al. "Genome-wide profiling of follicular lymphoma by array comparative genomic hybridization reveals prognostically significant DNA copy number imbalances," *The Blood Journal*, August 14, 2008; DOI 10.1182/blood-2008-02-140616 (43 pages).
- [5] Chi, B., deLeeuw, R., Coe, B., Ng, R., MacAulay, C. and Lam, W. "MD-SeeGH: a platform for integrative analysis of multi-dimensional genomic data," *BMC Bioinformatics*, 2008, 9:243doi:10.1186/1471-2105-9-243 (9 pages).
- [6] Lee, E., **Chari, R.**, Lam, A., Ng, R., Yee, J., English, J., Evans, K., MacAulay, C., Lam, S. and Lam, L. "Disruption of the non-canonical Wnt pathway in lung squamous cell carcinoma," *Clinical Medicine: Oncology*, 2008:2, pp. 169-179.
- [7] **Chari, R.**, Lonergan, K., Ng, R., MacAulay, C., Lam, W. and Lam, S., "Effect of Active Smoking on the Human Bronchial Epithelium Transcriptome," *BMC Genomics*, 8:297, pp. 1-13, August 2007.
- [8] **Cohen-Freue, G.**, Hollander, Z., Ng, R. et al. "MDQC: a new quality assessment method for microarrays based on quality control reports," *Bioinformatics Journal*, 23, 23, pp. 3162-3139, 2007.
- [9] **Shah, S.**, Lam, W., Ng, R. and Murphy, K., "Modelling Recurrent DNA Copy Number Alterations in array CGH Data," *Bioinformatics Journal*, 23, 13, pp. i450-i458, August 2007.
- [10] Kim, C., Shim, K. and Ng, R., "SQUIRE: Sequential Pattern Mining with Quantities," *International Journal of Systems and Software*, 80, 10, pp. 1726-1745, 2007.
- [11] Wong, K., Ng, R., Lam, W. et al. "A comprehensive analysis of common copy number variations in the human genome", *American Journal of Human Genetics*, 80, pp. 1-14, January 2007.
- [12] Ng, R. and Pei, J. Introduction to the special issue on data mining for health informatics. *SIGKDD Explorations*, 9, 1, pp. 1-2, 2007.
- [13] **Calders, T.**, Lakshmanan, L., Ng, R., "Expressive Power of an Algebra for Data Mining," *ACM Transactions on Database Systems*, 31, 4, pp. 1169-1214, December 2006.
- [14] **Chilson, J.**, Ng, R., Wagner, A. and Zamar, R., "Parallel Computation of High Dimensional Robust Correlation and Covariance Matrices," *the Journal of Algorithmica, Special Issue on Coarse Grained Parallel Algorithms For Scientific Applications*, 45, 3, pp. 403—431, July 2006.
- [15] **Malossini, A.**, Blanzieri, E. and Ng, R., "Detecting potential labeling errors in microarrays by data perturbation," *Bioinformatics Journal*, 22, 17, pp. 2214—2121, July 2006.
- [16] Lonergan, K., **Chari, R.**, Shadeo, A., deLeeuw, R., Chi, B., Tsao, M., Jones, S., Marra, M., Ng, R., MacAulay, C., Lam, S. and Lam, W., "Identification of novel lung genes in bronchial epithelium by serial analysis of gene expression," *American Journal of Respiratory Cell and Molecular Biology*, June 2006.

Raymond T. Ng (109895)

- [17] **Shah, S.**, Xuan, X., DeLeeuw, R., Khojasteh, M., Lam, W., Ng, R. and Murphy, K., "Integrating copy number polymorphisms into array CGH analysis using a robust HMM," *Bioinformatics Journal*, 22, 14, pp. 431—439, May 2006.
- [18] **Sander, J.**, Ng, R., **Sleumer, M.**, **Yuen, M.**, and Jones, S., "A Methodology for Analyzing SAGE Libraries For Cancer Profiling", *ACM Transactions on Information Systems, Special Issue on Genomic Information Retrieval*, 23, 1, pp. 35—60, 2005.
- [19] **Yanagawa, B.**, Taylor, L., Ng, R., Schereiner, G., Triche, T., Yang, D., and McManus, B., "Affymetrix oligonucleotide analysis of gene expression in the injured heart", *Methods in Molecular Medicine*, 112, pp. 305--320, 2005.
- [20] **Ying, A.**, Murphy, G., Ng, R., and Chu-Carroll, M., "Predicting Source Code Changes by Mining Change History", *IEEE Transactions on Software Engineering*, 30, 9, pp. 574--586, September, 2004.
- [21] Lakshmanan, L., **Leung, C.** and Ng, R. "Efficient Dynamic Mining of Constrained Frequent Sets," *ACM Transactions on Database Systems*, 28, 4, pp. 337-389, 2003.
- [22] **Kwon, A.**, Hoos, H., and Ng, R. "Inference of Transcriptional Regulation Relationships from Gene Expression Data," *the Bioinformatics Journal*, 19, 8, pp. 905-912, May 2003.
- [23] **Calders, T.**, Ng, R. and Wijzen, J., "Searching for Dependencies at Multiple Abstraction Levels", *ACM Transactions on Database Systems*, 27, 3, pp. 229-260, 2002.
- [24] Ng, R. and Han, J., "CLARANS: A Method for Clustering Objects for Spatial Data Mining", *IEEE Transactions on Knowledge & Data Engineering*, 14, 5, pp. 1003-1016, 2002.

B. Refereed Conference Publications (2002-2008)

- [25] Carenini, G., Ng, R. and **Zhou, X.** "Summarizing Emails with Conversational Cohesion and Subjectivity", *Proc. the 46th Annual Meeting of the Association for Computational Linguistics*, June 2008.
- [26] **Lee, H.**, Ng, R. and Shim, K., "Extending Q-grams to Estimate Selectivity of Strings with Low Edit Distance", *Proc. of 33rd International Conference on Very Large Databases*, pp. 195—206, September 2007.
- [27] Carenini, G., Ng, R. and **Zhou, X.** "Summarization of Email Conversations", *Proc. World Wide Web Conference*, pp. 91-100, May 2007.
- [28] **Gokhale, C.**, **Gupta, N.**, **Kumar, P.**, Lakshmanan, L., Ng, R. and **Prakash, A.**, "Complex Group-by Queries for XML," *Proc. 2007 International Conference on Data Engineering*, pp. 646-655, April 2007.
- [29] **Bu, S.**, Lakshmanan, L., Ng, R. and **Ramesh, G.**, "Preservation of Patterns and Input-Output Privacy," *Proc. 2007 International Conference on Data Engineering*, pp. 696-705, April 2007.
- [30] **Shah, S.**, Xuan, X., DeLeeuw, R., Khojasteh, M., Lam, W., Ng, R. and Murphy, K., "Integrating copy number polymorphisms into array CGH analysis using a robust HMM," *Proc. 14th International Conference on Intelligent Systems for Molecular Biology*, August 2006.
- [31] Carenini, G., Ng, R. and **Pauls, A.**, "Multi-document summarization of evaluative text," *Proc. 11th Conference of the European Chapter of the Association for Computational Linguistics*, April 2006.
- [32] Carenini, G., Ng, R. and **Pauls, A.**, "Interactive Multimedia Summaries of evaluative text," *Proc. 10th International Conference on Intelligent User Interfaces*, pp. 124-131, January 2006.
- [33] Carenini, G., Ng, R. and **Zwart, E.**, "Extracting Knowledge from Evaluative Text," *Proceedings of Third International Conference on Knowledge Capture*, pp. 11-18, October 2005.
- [34] **Bu, S.**, Lakshmanan, L. and Ng, R., "MDL Summarization with Holes," *Proceedings of 31st International Conference on Very Large Databases*, pp. 433-444, August 2005.
- [35] Carenini, G., Ng, R. and **Zhou, X.**, "Scalable Discovery of Hidden Emails from Large Folders," *Proceedings of ACM SIGKDD Conference*, pp. 544-549, August 2005.

Raymond T. Ng (109895)

- [36] Lakshmanan, L., Ng, R., and **Ramesh, G.**, "To Do or Not To Do: The Dilemma of Disclosing Anonymized Data," *Proceedings of ACM SIGMOD Conference*, pp. 61—72, June 2005.
- [37] **Chan, T.**, Ng, R., McCaulay, C., and Lam, W., "Finding Biomarkers Specific for Early and Late Stages of Lung Cancer Using SAGE Data", *Proceedings of 11th World Conference on Lung Cancer*, July 2005.
- [38] **Chari, R.**, Ng, R., McCaulay, C., Lam, S. and Lam, W., "Effect of Inhaled Budesonide in Smokers with Bronchial Dysplasia," to appear: *Proceedings of 11th World Conference on Lung Cancer*, July 2005.
- [39] Carenini, G., Ng, R., **Zhou, X.** and **Zwart, E.**, "Discovery and Regeneration of Hidden Emails", *Proceedings of the 2005 ACM Symposium on Applied Computing*, pp. 503--510, March 2005.
- [40] **Chen, L.** and Ng, R., "On The Marriage of Lp-norms and Edit Distance," *Proceedings of 30th International Conference on Very Large Databases*, pp 792--803, August 2004 (Acceptance rate: 81/504).
- [41] **Chilson, J.**, Ng, R., Wagner, A. and Zamar, R., "Parallel computation of high dimensional robust correlation and covariance matrices", *Proceedings of the ACM SIGKDD*, pp 533-538, August 2004.
- [42] Ng, R. and **Cai, Y.** "Indexing Spatio-temporal Trajectories using Chebyshev Polynomials", (**Best Paper Award**), *Proceedings of the ACM SIGMOD*, pp. 599--610, June 2004. (Acceptance rate: 69/431).
- [43] Jagadish, H., Ng, R., Ooi, B. and Tung, K., "ItCompress: an Iterative Semantic Compression Algorithm," *Proceedings of the 20th International Conference on Data Engineering*, pp. 646—657, March 2004.
- [44] **Kim, C., Lim, J.**, Ng, R. and Shim, K., "SQUIRE: Sequential Pattern Mining with Quantities," *Proceedings of the 20th International Conference on Data Engineering*, pp 827, March 2004.
- [45] **Chan, T.**, Ng, R., MacAulay, C., and Lam, W. "Using the Permutation Test to Analyze Lung Cancer SAGE Libraries", *Proceedings of the 10th World Conference on Lung Cancer*, August 2003.
- [46] **Chan, T.**, Ng, R., MacAulay, C., and Lam, W. "Evaluating Housekeeping Genes Used In Lung Microarray Analysis With Lung SAGE Data", *Proceedings of the 10th World Conference on Lung Cancer*, August 2003.
- [47] **Kwon, A.**, Hoos, H., and Ng, R. "Inference of Transcriptional Regulation Relationships from Gene Expression Data," *Proceedings of the 2003 ACM Symposium on Applied Computing*, pp. 135-140, 2003.
- [48] Ng, R., **Chan, T.**, and **Kuo, B.**, "Analyzing Brain and Breast SAGE Libraries", *Proceedings of the 10th International Conference on Intelligent Systems for Molecular Biology*, August 2002.
- [49] Ng, R., **Sleumer, M., Yuen, M., Sander, J.** and Jones, S., "Hierarchical Cluster Analysis and Classification of SAGE Data", *Proceedings of the 10th International Conference on Intelligent Systems for Molecular Biology*, August 2002.
- [50] Lakshmanan, L., Ng, R., **Wang, C., Zhou, X.** and Johnson, J. "The Generalized MDL Approach for Summarization", *Proceedings of 28th International Conference on Very Large Databases*, pp 766—777, August 2002 (Acceptance rate: 59/431).
- [51] **Leung, K.**, Ng, R. and Mannila, H. "OSSM: A Segmentation Approach to Optimize Frequency Counting", *Proceedings of International Conference on data engineering*, pp 583--592, March 2002 (Acceptance rate: 54/287).

3. Other Evidence of Contributions and Impact

A. Significant Applications and Translation

The Genome Canada funded project "Better Biomarkers in Transplantation" successfully completed earlier this year. Some of the findings are very promising. As a result, there will be a Canada-wide validation study involving hundreds of patients in Vancouver, Edmonton, Toronto and Montreal. When

Raymond T. Ng (109895)

this study completes in 2010, and if the results remain to be as strong as what we have witnessed in our discovery phase, the US Food and Drug Administration will likely fast-track our proposed biomarkers into the US market. We have just started our discussion with Health Canada.

B. Membership of Editorial and Advisory Boards

Associate Editor, the Very Large Database Journal, 2001-2007.

Associate Editor, the IEEE Transactions on Knowledge & Data Engineering, 2002-2008.

Member of the Research Management Committee, NCE IRIS, 2001-2003.

Member of the ACM SIGKDD Advisory Board, 1999—2001.

C. Keynote Plenary Presentations

“Discovering Outliers from Large Datasets”, Joint Statistical Meetings, August 1998, Dallas.

“Are Spatial Data Special: from a Data Mining Perspective?”, Third International Conference on Knowledge Discovery and Data Mining, August 1997, Los Angeles.

D. Program Chair or Co-chair

International Conference on Data Engineering, Program co-Chair, Shanghai, April 2009.

ACM SIGMOD General Chair, Vancouver, June 2008.

ACM SIGKDD Conference Program Co-chair, Edmonton, July 2002.

International Workshop on Data Mining and Discovery in Molecular Databases, Program Chair, January 1999.

SIGMOD Workshop on Research Issues on Data Mining & Knowledge Discovery, Program Co-chair, Montreal, June 1996.

E. Membership of Program Committees of Major Conferences

Conference on Very Large Data Bases (1997, 1998, 2001, 2004,2005,2007,2009).

ACM SIGMOD Conference(1999, 2003,2005, 2006).

ACM SIGKDD Conference (1996--2001, 2003,2004, 2005,2006,2007).

Conference on Data Engineering (2000, 2002, 2003,2005).

ACM Symposium on Principles of Database Systems (1997).

Pacific Symposium on Biocomputing (1998, 1999).

ACM GIS Conference (1998).

F. PhD External Examiner

Ji Zhang, Dalhousie University, 2008.

Toon Calders, University of Antwerp, Belgium, 2003.

Themis Palpanas, University of Toronto, 2003.

Vincent Cho, Hong Kong University of Science and Technology, Hong Kong, 2002.

Wu Lee, Carneige Mellon University, USA, 2002.

Markus Breunig, University of Munich, Germany, 2001.

Joerg Sander, University of Munich, Germany 1998.



**SEND ONE
ORIGINAL ONLY
DO NOT
PHOTOCOPY**

**APPENDIX A
Personal Data
(Form 100)**

Complete this appendix (i) if you are an applicant or co-applicant applying for the first time; (ii) if you need to update information submitted with a previous application; or (iii) if you do not hold an appointment at a Canadian postsecondary institution. For updates, include only the revised information in addition to the date, your name and your PIN.

This information will be used by NSERC primarily to contact applicants and award holders. It may also be used to identify prospective reviewers and committee members, and to generate statistics. It will not be seen or used in the adjudication process.

			Date 2008/10/20
Family name Ng	Given name Raymond	Initial(s) of all given names T.	Personal identification no. (PIN) 109895
Position and complete mailing address if your primary place of employment is not a Canadian postsecondary institution or if your current mailing address is temporary			If address is temporary, indicate: Starting date Leaving date
Telephone number (604) 822-2394	Facsimile number (604) 822-5485	E-mail address rng@cs.ubc.ca	
Telephone number (alternate)	<input type="checkbox"/> Give an alternate telephone number only if you can be reached at that number during business hours.		Gender (completion optional) <input checked="" type="checkbox"/> Male <input type="checkbox"/> Female
LANGUAGE CAPABILITY			
English	Read <input checked="" type="checkbox"/>	Write <input checked="" type="checkbox"/>	Speak <input checked="" type="checkbox"/>
French	Read <input type="checkbox"/>	Write <input type="checkbox"/>	Speak <input type="checkbox"/>
I wish to receive my correspondence:		in English <input checked="" type="checkbox"/>	in French <input type="checkbox"/>
AREA(S) OF EXPERTISE			
Provide a maximum of 10 key words that describe your area(s) of expertise. Use commas to separate them. If you have expertise with particular instruments and techniques, specify which one(s). data mining, data management, health informatics, text mining			Research subject code(s) Primary 2711 Secondary 2805



Appendix D (Form 100) Consent to Provide Limited Personal Information About Highly Qualified Personnel (HQP) to NSERC

NSERC applicants are required to describe their contributions to the training or supervision of highly qualified personnel (HQP) by providing certain details about the individuals they have trained or supervised during the six years prior to their current application. HQP information must be entered on the Personal Data Form (Form 100). This information includes the trainee's name, type of HQP training (e.g., undergraduate, master's, technical etc.) and status (completed, in-progress, incomplete), years supervised or co-supervised, title of the project or thesis, and the individual's present position.

Based on the federal *Privacy Act* rules governing the collection of personal information, applicants are asked to obtain consent from the individuals they have supervised before providing personal data about them to NSERC. In seeking this consent, the NSERC applicant must inform these individuals what data will be supplied, and assure them that it will only be used by NSERC for the purpose of assessing the applicant's contribution to HQP training. To reduce seeking consent for multiple applications, applicants will only need to seek consent one time for a six-year period. If the trainee provides consent by e-mail, the response must include confirmation that they have read and agree to the text of the consent form.

When consent cannot be obtained, applicants are asked to not provide names, or other combinations of data, that would identify those supervised. However, they may still provide the type of HQP training and status, years supervised or co-supervised, a general description of the project or thesis, and a general indication of the individual's present position if known.

An example of entering HQP information on Form 100 (with and without consent):

Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
Consent Received from Marie Roy				
Roy, Marie	Undergraduate (Completed)	Supervised 1994 - 1997	Isotope geochemistry in petroleum engineering	V-P (Research), Earth Analytics Inc., Calgary, Alberta
Consent Not Obtained from Marie Roy				
(name withheld)	Undergraduate (Completed)	Supervised 1994 - 1997	Isotope geochemistry	research executive in petroleum industry - western Canada

Consent Form

Name of Trainee	
Applicant Information	
Name Ng, Raymond T.	
Department Computer Science	Postsecondary Institution British Columbia
<p>I hereby allow the above-named applicant to include limited personal data about me in grant applications submitted for consideration to NSERC for the next six years. This limited data will only include my name, type of HQP training and status, years supervised or co-supervised, title of the project or thesis and, to the best of the applicant's knowledge, my position title and company or organization at the time the application is submitted. I understand that NSERC will protect this data in accordance with the <i>Privacy Act</i>, and that it will only be used in processes that assess the applicant's contributions to the training of highly qualified personnel (HQP), including confidential peer review.</p>	
_____	_____
Trainee's signature	Date
<p>Note: This form must be retained by the applicant and made available to NSERC upon request.</p>	