# White Paper on the Status of

# Statistical Science for 'omic Research in Canada

## February 25, 2008

## Preface

The workshop generating the ideas for this white paper was organized by Karen A. Kopciuk (Alberta Cancer Board/University of Calgary), Mayi Arcellana-Panlilio (University of Calgary), and Jennifer Bryan (University of British Columbia) as one of the activities of the Mini-Programme in Statistical Genomics. The organizers would like to recognize the assistance of Dr. Graham Chandler, professional writer, with the writing of this document. Generous funding for this mini-programme was obtained from the Pacific Institute for the Mathematical Sciences (PIMS) and for this workshop from the Banff International Research Station (BIRS) for Mathematical Discovery Innovation and Discovery and the Natural Sciences and Engineering Research Council of Canada. Support from Research Services at the University of Calgary is also gratefully acknowledged


# Statistical Science for 'omic Research in Canada

This paper is the product of a two-day workshop held at the Banff International Research Station in Banff, Alberta, Canada from June 29 to July 1, 2007. Eighteen of Canada's leading researchers (listed in Appendix 1) active at the interface of statistics and genomics were brought together in a series of facilitated sessions to explore challenges and opportunities facing 'omic research in Canada. This report summarizes those discussions and presents key recommendations.

## Executive Summary

Rapid advances in genome research over the past decade have engendered a unique and inventive interface between biology and statistics. It's a burgeoning field that has quickly grown into an entirely new specialty discipline tentatively called statistical genomics.

The rapid growth of statistical genomics has been driven by the vast amounts of data generated by today's genomics research that needs to ask increasingly complex questions in order to uncover relationships and interactions, and shed light on biology and meaning. Cutting-edge statistical research to develop novel analytical methods is now in great demand. It has become clear that genomic and statistical research must go hand in hand, each informing the other based on sound principles with the common goal of obtaining answers closer to the objective truth than either could have achieved alone.

Challenges have arisen in step with these advances. Funding agencies have not kept pace in recognizing that development of these critically required methodologies merits specific, independent financial support. University departments struggle to provide meaningful pathways for their graduate students and faculty in the new field. There is a strong push both from the statistical genomics discipline and the genetic research community for action on these issues.

To address these, four key recommendations emerged at the Banff workshop:

- Create new, stable sources of funding, or broaden existing sources to enable sustainable research and training environments;
- Support interdisciplinary training environments which can identify and select candidates with interdisciplinary potential and training in at least one key discipline, followed by provision of training in the remaining discipline(s) to the competence level required;
- Construct national or regional training centres in statistical genomics; and
- Develop strategies for statistical genomics researchers to fit and thrive in traditional, single discipline departments.

## Background

The past decade has seen rapid technological advances enabling the transition from reductionist single-gene studies to genome-wide approaches to understand biological phenomena. The sequencing of the human genome was an important first step, providing the basic information upon which questions of activity, expression, function and control might be addressed. Miniaturized high throughput technologies, such as DNA microarrays, have fueled genomic research, in this instance, by allowing the interrogation of thousands of genes in parallel. But new technologies generate comprehensive, high dimensional data very efficiently and have the potential of uncovering complex relationships. Researchers can now study a gene at various biological levels: expression (active or inactive genes), base pair changes within a gene (single nucleotide polymorphisms or SNPs), gene products (protein expression), and even metabolites.

Critically, efficient use of these new platforms is incumbent upon corresponding development of study designs, data management tools, methods of analysis and software. The experiments are fraught with multiple sources of variability so that findings can be difficult to repeat by other researchers. Beyond this basic fundamental of replicability is the necessity for increasingly sophisticated methods of design and analysis to address the complexity of the questions that need to be asked with these technologies.

Despite these challenges associated with new technologies, valid biological information has been rapidly accumulating. For example, the cystic fibrosis gene is now known to have more than 900 variations among its 250,000 base pairs; it can be modified by at least one other gene; and its protein structure and function are well documented. Searches for the myriad of complex relationships within these data will be enhanced as fresh and creative statistical techniques can be developed; speeding progress towards practical applications for preventing and treating the disease.

This recent 'omics revolution has fundamentally changed the way biologists conduct research. It is creating new research areas, and new opportunities for researchers from diverse backgrounds to work together on complex biological problems. It is fast-paced; new technologies are coming on stream at an increasing rate. Experiments utilizing the new technologies demand new study designs and methods of data analysis to untangle the biological process under study. Development of software designed for statisticians or biologists is an integral part of all proposed methods as it enables users to implement new and more appropriate methods. Integration of accumulating biological information requires special tools to link existing knowledge with relevant biology.

The rapidly emerging discipline of statistical genomics has become an integral component of genomic research and the need for new tools and practitioners continues to grow apace.

## Two success stories

Two examples of successful statistical genomic projects here illustrate the need for parallel development of advanced statistical methods with genomics and biology. They show how success can be achieved with the right levels and types of support.

### National Network in Statistical Genetic Modelling and Analysis of Complex Traits

The Network of Centres of Excellence (NCE) in Mathematics of Information Technology and Complex Systems (MITACS) currently funds a national multi-centre research project under the leadership of Dr. Shelley Bull of the University of Toronto. The research conducted by members of the project team (predominantly statisticians and quantitative genetic epidemiologists) focuses on problems posed by local scientific collaborators in human and molecular genetics, population genetics, forestry genetics, cancer biology, and genetic epidemiology. The project team is a network of investigators at six universities across Canada, developing theoretical and computational solutions to specific analytic questions arising in collaborations and in the field in general. As detailed at the website, (http://www.mshri.on.ca/mitacs/publications/publications_presentations.html)

these include published manuscripts in theory and methods. In addition, methods are disseminated in software implementing novel computational methods.

Productive and ongoing collaborations have been established with partners affiliated with industry, biomedical research institutes, hospitals, and research networks and foundations. In this interdisciplinary environment, the partners educate project team members and their trainees about their field and provide funds for complex data analysis and methods implementation. Matching mechanisms generate additional funds for more fundamental statistical methods development and evaluation, typically through trainee research projects. The trainees usually have strong quantitative or statistical theory backgrounds from formal training in statistics, and are able to acquire experience both in statistical research and in complex applied data analysis, as well as develop their biology background. Expertise in study design and in newly-developed and existing genetic analysis approaches is applied to scientific investigations. These include statistical genetic study design and analysis of data in: cystic fibrosis, respiratory diseases, inflammatory bowel disease, kidney disease, hypertension, diabetes, cancers, infection, schizophrenia, kidney stones, mouse models, and tree diversity. Part of this team's mandate is to bring new ideas, methods, and techniques from mainstream statistics to bear on analytic issues and problems relevant to research in statistical genetics and genomics.

Importantly, the project has enabled trainees to make the transition into statistical genomics. An example: "Working under the leadership of Dr. Shelley Bull at the University of Toronto, I was immersed in a talented research group that has both breadth and depth in this challenging research area. The opportunities to learn, share and discuss ideas were numerous, from journal clubs and group meetings to seminars and discussions with basic and clinical scientists. Collaborations led to a successful CIHR operating

grant, ideas for my first NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Award and a new research program," writes Karen A. Kopciuk, in *MITACS Connections,* December 2003.

As a group, team members have been responsible for highly qualified people (HQP) training in genetic/genomic statistical methods across Canada. The focus has been on individuals with strong mathematical, statistical, or computational backgrounds, providing research training in interdisciplinary, biologically and clinically oriented settings with opportunities to work with investigators in various disciplines. Trainees include post-doctoral fellows, PhD and MSc students, undergraduate co-op and summer students, and research assistants. All former trainees are now employed in the biomedical sector, including research institutes and pharmaceutical companies, or in research universities. (http://www.mshri.on.ca/mitacs/members/students.html) This cohort of researchers represents a major achievement in expanding the intellectual capital in genetic and genomic statistical methods in Canada.

## Importance of national funding for statistical methods research

In 2003, Brad McNeney and Jinko Graham of Simon Fraser University received a three-year methods grant from the Canadian Institutes of Health Research (CIHR) Institute of Genetics Request for Applications (RFA) on Novel Population Genetic and Genetic Epidemiological Methods for Studies of Complex Genetic Diseases. The focus of their proposed research was statistical methods for simultaneous estimation of haplotypes and haplotype versus environmental risks in association studies of complex genetic diseases. The funding improved their competitive position for a 2004 RFA from the CIHR Institute of Population and Public Health on Advancing Theories, Frameworks, Methods and Measurement in Health Services and Policy, Population and Public Health Research and Knowledge Translation.

Importantly, the 2004 operating grant provided critical teaching buyout for McNeney and both grants were instrumental in securing a salary award with teaching release for Graham through the BC Michael Smith Foundation for Health Research. Without the funding for release time, it would have been difficult to establish and maintain research productivity and collaborations with medically-oriented colleagues at other institutions. Benefits have been far reaching: the statistical research resulting from these opportunities has since motivated and been applied by collaborators studying diabetes (e.g. Bekris *et al*. 2007, Shin *et al*. 2007,), asthma and chronic obstructive pulmonary disease (He *et al*. 2006, He *et al*. 2007, Tanaka *et al*. 2007, Wallace *et al*. 2006, Zhang *et al*. 2007) and non-Hodgkin lymphoma (Novik *et al*. 2007). In addition, four freely-available software packages (hapassoc, LDheatmap, LUCA and elrm) have been implemented in R, which can be downloaded from the Comprehensive R Archive Network (http://cran.stat.sfu.ca).

Furthermore, several associated papers (Burkett *et al*. 2006, Shin *et al*. 2006, Shin *et al*. 2007, Zamar *et al*. 2007) have been published. Their group also participated in the 14th Genetic Analysis Workshop (Burkett *et al*. 2005), a collaborative effort of the genetic epidemiology community with a significant training component for evaluating and

comparing novel methods. Also seven MSc students have been trained to date. Much of this research and training would not have occurred without the crucial initial support from CIHR for statistical methods research and training.

It is seen that initial funding is critical in laying a foundation for Canadian research at the interface of genetics, statistics and epidemiology. We believe that strengthening the methodological components of such interdisciplinary collaboration will lead to important synergies and take Canadian genetics research to the next level. The emerging Canadian community of methodological researchers has much potential and is eager to become more involved.

## Challenges limiting success

Key features that enabled success can be drawn from these two cases. Where these features are lacking, challenges arise which will substantially limit success. Those challenges are grouped into three important categories.

### Challenge 1: Funding

Currently, funding is sourced primarily from NSERC Discovery Grants, our largest source of research funds. During the 2007 grant application year, 106 grants were awarded to researchers by the Statistical Science Grant Selection Committee (GSC 14). The success rate for Discovery Grants in the discipline has been high (71% in the most recent competition, see Appendix 2) and the application process is relatively straightforward. Their effectiveness, however, is limited by the small amounts of individual grants: the median amount in the 2007 competition was $17,100 per year. This level is simply too small to be our sole source—it's typically sufficient to fund at most one graduate student and does not permit hiring of support staff such as programmers or data base managers.

Cross-disciplinary and/or small-team grants are often a good fit, especially with respect to amounts, team size and team composition. NSERC/CIHR Collaborative Health Research Projects (CHRP) would be ideal to fund mid-sized teams of two to four investigators, encompassing experiments and the associated analysis. The additional requirement for a concrete and near-term application associated with these grants, however, can be difficult to meet given the nature of statistical genomics research.

Requirements of larger science organizations such as Genome Canada or MITACS pose other challenges. They require prospects of immediate application to human health, commercialization, or matching funds from a corporation or entity. Moreover, MITACS projects and similar large scale endeavours require established researchers with the skills, contacts and experience to conduct them successfully. Being a young discipline, there are few statistical genomics researchers so qualified. Thus, despite its well-documented

successes, this type of large scale funding mechanism in the short term is not a universal solution.

Whereas CIHR and other agencies' RFAs may be the right size, they are problematic in many ways. Even if applications are successful, the one-off nature of these opportunities means they are not the reliable source of recurring operating funds our new discipline requires. Renewal or revised applications can be submitted to the regular operating grants panel within CIHR or another agency, but without a committee dedicated to statistical methodology and theory, success with obtaining CIHR Operating Grants is contingent on knowledgeable individuals with appropriate expertise being available and willing to serve on committees. The reception of bioinformatics proposals differs substantially across panels in the CIHR and seems to be poorest for the genomics panel. Some positive signs are however evident: the Genetics Peer Review Committee funded two statistical genomics projects in the March 2007 operating grants competition (submitted by investigators who had previously been funded by the strategic RFA in Novel Methods).

Another reason for the lack of success may be the disconnection between what funding agencies such as CIHR deem necessary (and therefore fund-worthy) research resources and what the field of statistical genomics requires. Our funding requests are typically dominated by costs for (a) graduate students, (b) postdoctoral students, and (c) other personnel, such as MSc statisticians performing analyses or creating statistical software, because the training of students and post-docs is integral to undertaking statistical genomics research. However, while CIHR Operating Grants allow funding of doctoral and post-doctoral trainees on projects, their primary purpose is not to fund trainees—the expectation has been that trainees should be able to source their own funding in many cases. Cobbling together individual trainee grants may serve, but they are not sufficient to maintain a statistical genomics research lab, and cannot be construed as a long-term solution. It fails to recognize that statistical genomics requires the integration of research training as much as other research programs require tubes, pipettes and centrifuges.

A further problem involves the shortage of peer representation on committees, which has the unfortunate result of making our proposals less understood, less appreciated and less championed. Other factors have contributed to low success rates. Mathematics and statistics departments, often the source of statistical genomics researchers, do not typically require post-doctoral training for academic appointments. Compared to biological scientists with the same academic rank (particularly at the Assistant Professor level), statisticians in the field generally have less experience, fewer publications, and hence are less competitive for external funding as the lead investigator.

Even at the administrative level, there is potential for misunderstanding between NSERC and CIHR over which agency has the mandate to support statistical genetics and genomics. For example, CIHR staff recently encouraged organizers of an interdisciplinary workshop with a substantial statistical methods component to enquire about funding from NSERC. The enquiry was promptly directed back to CIHR by NSERC staff. Such lack of ownership by any agency further confounds our funding picture.

Clearly, mechanisms for funding basic research that is interdisciplinary and collaborative are eluding us. Either analysis and methodology needs to be funded on its own or mechanisms need to be created to jointly fund the creation and application of statistical methodology. How can we address this? What changes in the present systems of research funding would help us? What do we need to do?

For starters, larger NSERC Discovery Grants are imperative. One approach, particularly if total funding levels remain static, is to change the panel's tradition of granting small individual amounts to large numbers of researchers. This is happening already. NSERC recently instituted a minimum grant award of $15,000, so with no infusion of new funds, fewer awards will be forthcoming. Unfortunately however, this new funding strategy risks penalizing new researchers and those working in emerging areas. Time will tell if such revisions lead to adequate funding for interdisciplinary research.

To enhance our chances of success at CIHR, review panels must have members with expertise in statistics and statistical genomics. We need to submit lists of potential reviewers to relevant panels such as Genetics, Genomics and Population & Public Health. Statistics representation on these panels would raise awareness and educate fellow panel members about critical aspects of statistical genomics research.

We need the capability to compete for renewable grants for statistical genomics research instead of the more typical one-off RFAs. The best solution would be for CIHR to establish an operating grant home specifically for statistical genomics researchers. It could take the form of a CHRP-like program (probably NSERC/CIHR) but without the requirement for immediate application in human health or commercialization and with the possibility of renewal. It should consist of two PIs, two budgets and a shared mission. As well, CHRP should be extended to promote biostatistics projects of $150K to $250K per year with two or three investigators (one statistician and one or two bioscience researchers). The recent doubling of available funds for the next five years in the CHRP funding program will go a long way towards such a structure. The objectives of the program have not changed. They still focus on applications leading to health benefits for Canadians, including improved health services or possible economic benefits from health-related technologies.

Because the new discipline is growing at an exponential rate, securing appropriate funding is imperative. These strategies to improve the funding environment are needed now.

## Challenge 2: Training and Capacity Building

We have an urgent need to generate opportunities to grow the ranks of senior investigators in statistical genomics. One basic challenge is that we do not yet have an effective and instantly-recognizable brand name, one which highlights ourselves as a distinct discipline and creates an awareness of potential career paths and how to embark on them.

Although it would be ideal, we don't really need students to have backgrounds in both statistics and biology. A better plan would be to accept students based on their excellence in one core discipline and their potential to acquire competence in the second discipline, and to think and work creatively to marry the two. We need to get better at finding, creating and nurturing people with these dual capacities who can develop strong collaborations.

There are many ways to accomplish this. Several large US centres offer top candidates guaranteed funding along with little or no graduate assistantship teaching load for several years. For instance, the Genomic Sciences Training Programs at the University of Wisconsin-Madison and the University of Michigan support pre-doctoral, postdoctoral and research trainees through grants from the National Human Genome Research Institute. Support includes tuition, stipends and travel funds. One Canadian success story is the University of British Columbia's CIHR Strategic Training Grant for Bioinformatics, which cross-trains biologists and computer science students. These and similar models might be looked at for statistical genomics training.

A further challenge is in the research environment. Where there is only one or but a few of us at an institution, that isolated environment makes it difficult to attract the strongest students and to provide them the best training. To address this, we should share the burden of training with local collaborators and statistical colleagues across Canada. A grander scheme, and an attractive one, would be for the discipline to band together and create one Canada-wide training program in statistical genomics. It would be a massive undertaking requiring a strong application, but the potential payoff is high. A somewhat less ambitious, but entirely valuable, alternative would be to develop regional programs.

Whether national or regional, such training programs would see an alliance of several institutions. Graduate students would take core and elective courses, be jointly supervised, rotate through different labs, and attend periodic workshops, such as at BIRS. A steering committee would set the curriculum, develop policies and procedures, and strategize for further development and sustainability. The CIHR would be one natural funding source for such a proposal as one of their strategic training programs. These national or regional training schemes however are at present a longer term vision, as our new discipline currently lacks the requisite cadre of senior people, and will need many years to develop and get into operation.

These schemes would provide a better flow of top students into the field. To ensure they stay, steps must also be taken to ensure they can look forward to long and fulfilling careers. Prospects of permanent positions in research institutions and universities will help. The current environment of soft funded positions following doctoral or post-doctoral training provides a dim outlook for students considering training beyond a master's degree. This leads to our next and final challenge.

## Challenge 3: Faculty recruitment, development and retention

Currently, few Canadian institutions have the critical mass in statistical genomics. The majority of researchers in our new discipline are embedded in traditional mathematics and statistics departments with some joint appointments in a department such as Medicine or health research agencies.

With existing departmental structures and boundaries, it can be difficult to mesh cross-disciplines within conventional academic systems. Departments want to boast of having biostatistical expertise, but often do not understand the need to expend departmental resources to facilitate its development. Holding joint appointments has a real disadvantage of our being evaluated on multiple, sometimes conflicting, criteria or on criteria not appropriate to our research.

As members of a new discipline with few faculty positions, we are more likely to get involved in grant reviews, refereeing, writing team grant applications, and serving on supervisory committees outside the department, than do our more classical colleagues at the same level of rank and expertise. These activities require significant time and energy, and need to be recognized.

Moreover, our discipline's collaborative nature can complicate individual career development. Forming and nurturing the interactions necessary to our research is time-consuming and not always reflected as measurable productivity. Since we collaborate closely with biological and medically-oriented disciplines, we often publish in journals unfamiliar to our statistics colleagues. Examples of these include *Genetic Epidemiology, Human Heredity*, and *BMC Genetics*. And, our research output may be in the form of software, which is not recognized on the same level as are conventional academic publications. These would be published in journals such as *Journal of Statistical Software* and *BMC Bioinformatics*.

We need to encourage academic departments to give credit for the work of interdisciplinary faculty that may not fit into traditional measures of productivity. An important way to achieve this is by educating our colleagues. Visibility is also paramount: we need to make ourselves known at meetings and conferences related to our research areas.

Our discipline needs the recognition of an appointment such as a Canada Research Chair in Statistical Genomics. This would allow a statistical genomics researcher to work as an independent investigator in a collaborative environment, the very conditions required for such research to flourish. These research chairs would be tenable in institutions across the country and naturally enable linkages, the flow and exchange of information and expertise, thereby strengthening the interdisciplinary research needed to move the field forward.

Teaching release, using small amounts of funds from operating grants, would have a large impact. In more service-orientated departments, such as Mathematics and Statistics,

the undergraduate teaching load can be substantially heavier than in lab-sciences departments. These heavier teaching loads hinder collaborative research activities throughout most of the academic calendar year. Being able to use operating grants to buy out teaching commitments would be another way small changes in the policies of agencies such as CIHR could make a difference.

## Conclusions

Rapid development of new high throughput technologies and accumulation of complex biological information from different perspectives are demanding ever more powerful and sophisticated statistical methods, creating the new specialized discipline of statistical genomics. Banff workshop participants identified a number of challenging problems that face the new discipline; that require novel and complex solutions.

Potential solutions for three primary challenges have been identified. Firstly, it was agreed that statistical genomic research does not fit the mandate of most genomic research funding agencies and generally, traditional statistical funding agencies are not providing adequate funds to carry out necessary empirical validations. Strategies to improve the funding environment have been suggested. Although there have been some positive signs, implementation of these strategies is needed now.

Secondly, the workshop evaluated problems and suggested solutions for creating a workable training and career environment that will attract and retain new members in the discipline. These include the need to develop collaborations within and amongst universities and to provide training opportunities with adequate time for example by off-loading some teaching responsibilities during training. National networks and strategies to strengthen interdisciplinary research were suggested as requirements to advance the field; as well as making a natural home for it within university departmental structures.

Finally, promising career paths are needed to generate the critical mass necessary to grow and sustain our new discipline in Canadian universities and institutions. The current situation has failed to attract large numbers into the field, even for those with masters degrees in statistics with an emphasis on genetics or genomics. To remedy this, we have made recommendations ranging from reduction of undergraduate teaching loads to establishment of a Canada Research Chair in Statistical Genomics.

## References

Bekris L, Shepard C, Janer M, Graham J, McNeney B, Shin J-H, Zarghami M, Griffith W, Farin F, Kavanagh T, Lernmark A. Glutamate cysteine ligase catalytic subunit promoter polymorphisms and associations with type 1 diabetes age-at-onset and GAD65 autoantibody levels. Exp Clin Endocrinol Diabetes 2007 Apr;115(4):221-8. PubMed ID: 17479437

Shin J-H, Janer M, McNeney B, Blay S, Deutsch K, Lernmark A and Graham J for the Swedish Childhood Diabetes Study Group and the Diabetes Incidence in Sweden Study Group. IA-2 autoantibodies in incident type I diabetes patients are associated with a polyadenylation signal polymorphism in GIMAP5. Genes and Immunity 2007 Sept;8(6):503512. PubMed ID: 17641683

He JQ, Burkett K, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Interferon gamma polymorphisms and their interaction with smoking are associated with lung function. Human Genetics 2006 119(4):365-375

He JQ, Shumansky K, Zhang X, Connett JE, Anthonisen NR, Sandford AJ. Polymorphisms of interleukin-10 and its receptor and lung function in COPD. European Respiratory Journal 2007 29(6):1120-1126

Tanaka G, Sandford AJ, Burkett K, Connett JE, Anthonisen NR, Pare PD, He JQ. Tumour necrosis factor and lymphotoxin A polymorphisms and lung function in smokers European Respiratory Journal. 2007 29(1):34-41

Wallace AM, He JQ, Burkett KM, Ruan J, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Contribution of alpha- and beta-defensins to lung function decline and infection in smokers: an association study. Respiratory Research 2006 7: Article 76

Zhang XZ, Mahmudi-Azer S, Connett JE, Anthonisen NR, He JQ, Pare PD, Sandford, AJ. Association of Hck genetic polymorphisms with gene expression and COPD. Human Genetics 2007 120(5):681-690

Novik KL, Spinelli JJ, MacArthur AC, Shumansky K, Sipahimalani P, Leach S, Lai A, Connors JM, Gascoyne RD, Gallagher RP, Brooks-Wilson AR Genetic variation in H2AFX contributes to risk of non-Hodgkin lymphoma. Cancer Epidemiology Biomarkers and Prevention 2007 16(6):1098-1106

Burkett K, Graham J, McNeney B. hapassoc: Software for likelihood inference of trait associations with SNP haplotypes and other attributes. Journal of Statistical Software 2006 16(2):1-19.

Shin J-H, Blay S, McNeney B, Graham J. LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. Journal of Statistical Software 2006, 16 Code Snippet 3.

Shin J-H, McNeney B, Graham J. "Case-Control Inference of Interaction between Genetic and Nongenetic Risk Factors under Assumptions on Their Distribution," Statistical Applications in Genetics and Molecular Biology 2007 6(1), Article 13. Available at: http://www.bepress.com/sagmb/vol6/iss1/art13

Zamar D, McNeney B and Graham J. elrm: Software implementing exact-like inference for logistic regression models. Journal of Statistical Software 2007 21(3):1-18.

Burkett K, Ghadessi M, McNeney B, Graham J and Daley D. A comparison of three methods for selecting tagging SNPs. BMC Genetics 2005, 6:S71.

*Appendix 1*

# List of Participants attending the Statistical Science for `omic research in Canada at BIRS, June 29 - July 1, 2007

| Name | Affiliation |
|---|---|
| Arcellana-Panlilio, Mayi* | University of Calgary |
| Brettschneider, Julia | Queen's University |
| Briollais, Laurent | Mount Sinai Hospital |
| Bryan, Jennifer* | University of British Columbia |
| Bull, Shelley | University of Toronto |
| Chandler, Graham | Independent Writer |
| Chen, Jiahua | University of British Columbia |
| Gottardo, Raphael | University of British Columbia |
| Graham, Jinko | Simon Fraser University |
| He, Wenqing | University of Western Ontario |
| Kopciuk, Karen* | University of Calgary/Alberta Cancer Board |
| Lesperance, Mary | University of Victoria |
| McNemey, Brad | Simon Fraser University |
| Nadon, Robert | McGill University |
| Ouellette, Francis | Ontario Institute for Cancer Research |
| Stephens, David | McGill University |
| Surette, Michael | University of Calgary |
| Turinsky, Andrei | University of Calgary |
| Wasserman, Wyeth | University of British Columbia |
| West, Sherry | University of Calgary |

*organizer

## *Appendix 2*

NSERC 2007 GSC 14 statistics

Source:
Judie Foster
Program Officer/Administratrice de Programme
Discovery Grants/Subventions à la découverte
NSERC / CRSNG

| **Application type** | **Number applied** | **Number successful (%)** | **Average annual grant awarded** |
|---|---|---|---|
| New | 28 | 20 (71%) | $14,900 |
| Renewal | 78 | 55 (71%) | $18,700 |
| | | | |
| Combined | 106 | 75 (71%) | $17,100 |
| | | | |