POLICY EVALUATION:
A CASE STUDY OF GENOME CANADA PROGRAMMING 2000-2011

A Thesis Submitted to the

College of Graduate Studies and Research

In Partial Fulfillment of the Requirement

For the Degree of Master of Arts in Public Policy

In Johnson-Shoyama Graduate School of Public Policy

University of Saskatchewan

By Lucy Chen ZHANG

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Graduate degree from the University of Saskatchewan, I agree that the Libraries of the University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College of Graduate Studies and Research. It is understood that any copying or publication or use of this thesis or parts for financial gain shall not be allowed without my permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or in part shall be addressed to:

Head of Johnson-Shoyama Graduate School of Public Policy

University of Saskatchewan

Saskatoon, Saskatchewan

S7N 5B8

ABSTRACT

Zhang, Lucy Chen, Master of Public Policy, Johnson-Shoyama Graduate School of Public Policy, University of Saskatchewan, Canada, 2013.

Policy Evaluation: A Case Study of Genome Canada Programming, 2000-2011.

Supervisor: Dr. Peter W. B. Phillips

The policy evaluation literature on research programing generally focuses on the cost-benefit of different choices in research systems. This thesis applies evaluation tools to assess the fit between project allocations and the strategic goals of Genome Canada, a major research funding organization in Canada.

Genome Canada (GC) was established in April, 2000, to provide funding and information resources related to genomics research. The research targets many key areas, such as health, agriculture, environment, forestry, energy, mining and fisheries.

Since then the scientific community has partnered with government, the private sector, and international organizations to fund research projects on genomics related subjects. Four open competitions (I, II, III and Applied Genomics in Bio-products and Crops or ABC), combined with a wide array of more targeted projects, have collectively been allocated more than C$2 billion in total investment for the 2000-2014 period.

This study assesses how well these research projects fit the stated goals of Genome Canada. The study assesses the fit between the goals and research investment decisions of GC. As a first step in this research, we conducted a review of Genome Canada operations to develop the background understanding of the system and its structure. After reviewing the goals, structure, selection processes and progress reports, we found that there was no explicit assessment of the fit between the stated goals and resource allocation decisions. This study targets to fill this area.

Second, we investigated the methods used by GC to develop and implement their goals. Once we understood these methods, we developed a research approach to assess

the fit between the goals and the outputs. The model was built to test each project against the stated overall program objectives, namely to: develop and implement a coordinated strategy for the technology in Canada; bring together industry, governments, universities, research hospitals and the public to support large-scale genomics and proteomics research projects; provide accessibility to science & technology platforms to researchers; and assist in attracting co-funding for projects from both domestic and international investors.

Third, we determined that the review processes contain scientific, financial and management criteria. By using the STATA tool, we tested the relationship between the stated goals of the organization and the share of funds allocated to specific projects both in the total pool of investments and the open competitions.

The analysis revealed that the overall fit for the entire investment program between 2001 and 2011 was about 35%, which is quite reasonable for such an analysis. We found the most important variable affecting resource allocation was the quality of the principal investigator. Other stated goals of GC were either less important or insignificant. By segmenting the analysis into the open-competition investments alone, we discovered the fit deteriorated (R^2 of 34% dropped to 22%), which suggests the directed investments are a stronger fit with the goals. While we could not conclusively determine the cause, it might be attributed to either weaknesses in the competitive process or a particularly effective and strategic effort by Genome Canada staff. Further analysis would be needed to determine this.

KEY WORDS: evaluation; research management; Genome Canada; program assessment.

ACKNOWLEDGEMENTS

Completion of this thesis would not have been possible without the support and encouragement of many people.

I would like to gratefully thank the faculty and staff of Johnson-Shoyama Graduate School of Public Policy. Their attitude towards the work encourages me to work harder and face the difficulties on my research road.

I would particularly like to thank my supervisor, Peter W.B. Phillips, for not only his deep thinking and variety of knowledge through the whole research, but also his gentle style and being truly a soul mentor during the study.

Thank you specifically to Heather McWhinney for introducing me to the art of English communication. The effort spent on words and sentences has been a great gift which was very helpful in completing the study.

As well, I would like to acknowledge the thoughtful guidance of Haizhen Mou who taught me STATA tool as the statistics method to solve the problems.

Finally, I would like to thank my mom and dad for their emotional support—they let me know they will always being on the side no matter what happened. A hard-working time will make you stronger and benefit all lifelong.

TABLE OF CONTENTS

1. Introduction

1.1. Genome Canada Overview

Genome Canada (GC) was established in April, 2000. It is an example of an independent non-profit organization, that provides funding, coordination and information resources related to research, in its case for genomics and proteomics research in Canada. The research targets the development and implementation of strategies and large scale research projects in key bio-science areas (i.e., health, agriculture, environment, forestry, fisheries, mining and energy) in order to help Canada become a world leader in genomics and proteomics research.

Genome Canada is based on the premise that the funding and management of large-scale interdisciplinary and internationally peer-reviewed research projects along with S&T (science and technology) Innovation Centers can effectively translate research results into broader commercial outcomes. Genome Canada operates in close collaboration with its primary partners—the six Genome Centers, located in British Columbia, Alberta, the Prairies, Ontario, Quebec, and the Atlantic region. The relationship established between Genome Canada and each of the Genome Centers is defined by means of a funding agreement that "not only acknowledges the independence of each Genome Centre, but also specifies the parameters in which each Centre is to operate and contribute to Genome Canada's overall mandate".[1]

Over the past decade, Genome Canada has established Canada as a recognized world leader in the promotion of research on the ethical, environmental, economic, legal and social (GE3LS) aspects of genomics research. While GC has undergone the usual organizational, administrative and financial reviews, it has not undertaken any specific evaluation of the process of targeting its operating model to realize its stated goals. This thesis addresses that gap.

1.2. **Problem statement**

A critical part of any effective public policy assessment is to compare activities and outputs against the authorized goals and objectives of the initiative. In most cases, the outputs are assumed to conform to the stated goals and objectives but are not assessed as part of a formal evaluation.

This project explicitly assesses the choices made by Genome Canada in the context of its funding competitions to determine how the organizational goals are reflected in the projects selected.

1.3. **Approach**

Genome Canada has developed a detailed operational style. While the order of the early steps in each competition might vary, all of the competitions followed a common path.

First, after consultation with industry, government, the scientific community and end users, (sometimes informally and sometimes through the use of formally structured theme papers), GC would frame a funding request for Industry Canada that states what area the organization would focus on and what the money would be used for. If successful, GC would then devise competition objectives. Most federal requests were only partially awarded.

Second, GC would issue a call for proposals, which would articulate the focus and scale of projects that could be funded. In most cases letters of intent are first reviewed and in a few cases were used to triage the proposals. Projects would be evaluated and invited to submit full proposals. Full proposals for the open competitions would be peer-reviewed and assessed by panels of international reviewers. The Genome Canada Board would then approve the allocations. Each approved project embodied milestones which would trigger quarterly progress reports and a final statement of activities and outputs.

In the context of this effort, Genome Canada regularly undertakes financial reporting that is audited, has engaged in organization and process evaluations and has assessed the outputs of the competitions. To date, the organization has not obviously assessed the

efficacy and appropriateness of the funding allocation decisions themselves and their fit to the organization's mandate and objectives.

1.4. **Structure**

Our study is designed to assess how well Genome Canada's funding allocations fit the organization's stated goals. This work is structured into five further chapters. Chapter 2 offers an overview of past GC reports and budgets to provide a background to the funding issues and models. Chapter 3 reviews the literature and theory of evaluation relevant to this work. Chapter 4 lays out the research method we use to examine the fit between the goals and the funding allocation decisions. Chapter five presents the results of our analysis. Chapter six examines the policy implications of this study.

2. Background

Genome Canada is a not-for-profit non-government-controlled organization set up by the federal government to invest in genomics research in key sectors and foster networks of expertise in Canada with a view to generating economic and social benefits for Canadians.

Over the past decade, Genome Canada has established Canada as a recognized world leader in genomics research. The unique approach Genome Canada has adopted ensures GE^{3}LS (the ethical, environmental, economic, legal and social) aspects are considered and integrated into science-based genomics research and large-scale research projects. This is posited to have helped enable responsible and beneficial applications of genomics science.

2.1. **Objectives of Genome Canada**

Genome Canada identified five key objectives to help move Canada onto the world stage in its 2007 corporate strategic plan.[2] Specifically, the organized seeks to:

1) Develop and implement a coordinated strategy for genomics and proteomics research to enable Canada to be among the world leaders.

2) Support large-scale genomics and proteomics research projects of strategic importance to Canada, which are beyond current capacities, by bringing together industry, governments, universities, research hospitals and the public.

3) Provide accessibility to Science & Technology Platforms to researchers in all genomics and proteomics related areas through six regional Genome Centers across Canada (Atlantic, Québec, Ontario, Prairie, Alberta and British Columbia). The relationship established between Genome Canada and each of the Genome Centers is defined by means of a funding agreement that not only acknowledges the independence of each Genome Centre, but also specifies the parameters in which each Centre is to operate and contribute to Genome Canada's overall mandate.

4) Encourage investment by others in the fields of genomics and proteomics, attracting co-funding for projects from both domestic and international investors.

5) Sustain leadership in research areas on Ethical, Environmental, Economic, Legal and Social issues related to genomics and proteomics research (GE^3LS), and promote the communication of the relative risks, rewards and successes of genomics and proteomics research to the Canadian public.

2.2. Genome Canada Governance System

Genome Canada operates within a governance framework that is reflective of its not-for profit corporation status. It is governed by a Board of Directors comprising up to 16 individuals drawn from the academic, private and public sectors. These individuals bring unique skills and experiences as well as strong interests and insights to successfully fulfill Genome Canada's mandate.

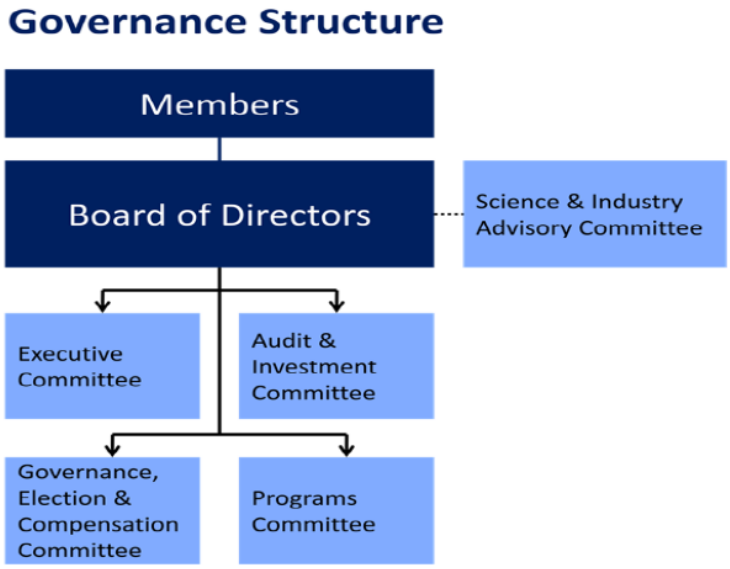


Figure 2.1 Genome Canada Structure

Source: Performance, Audit and Evaluation Strategy 2012-2017[3]

2.3. Funding & Investments

As can be seen in table 2.2, as of 2012 Genome Canada had committed $915 million in funding and researchers had secured approximately an additional $1085 million in co-

funding, representing a total investment of over $2 billion in completed or planned genomics research in Canada.

All these investments have laid a foundation for a rich, vibrant genomics research community in Canada, and as noted below, have transformed the quantity, scope, scale and quality of such research. [4]

Table 2.2 Operating Budgets

Details *(In millions of dollars)*	Forecast Cumulative 2000–01 to 2010–11	Planned 2011–12	Planned 2012–13 to 2013–14	Forecast Cumulative 2000–01 to 2014-15	Estimated Co-funding From 2000–01 to 2014–15	Total Genome Canada & Co-funding	Percent-age %
RECEIPTS							
Government of Canada	766.0	52.2	96.8	915.0		915.0	*43.8*
Investment Income	86.4	0.5	0.8	87.7		87.7	*4.2*
Co-Funding					1,085.4	1,085.4	*52.0*
	852.4	**52.7**	**97.6**	**1,002.7**	**1,085.4**	**2,088.1**	*100.0*
PROGRAM DISBURSEMENTS							
Research Projects							
Competition I	80.6			80.6	74.0	154.6	*7.4*
Competition II	146.2			146.2	137.6	283.8	*13.6*
Competition III	205.9			205.9	221.0	426.9	*20.5*
Multi-Sector Competition		10.0	20.0	30.0	30.0	60.0	*2.9*
Forestry and Environment		10.0	20.0	30.0	30.0	60.0	*2.9*
Applied Genomics In Human Health Competition	59.9			59.9	71.4	131.3	*6.3*
Applied Genomics in Bioproducts and Crops	16.7	15.9	22.4	55.0	59.0	114.0	*5.5*
Bovine Genome Sequencing Project	6.0			6.0	63.4	69.4	*3.3*
Structural Genomics Consortium	31.4	0.9		32.3	157.4	189.7	*9.1*
Public Population Project in Genomics	15.8			15.8	38.9	54.7	*2.6*
International Regulome Consortium	2.6			2.6	0.4	3.0	*0.1*
International Barcode of Life	1.3	5.4		6.7	6.7	13.4	*0.6*
Genome Canada-Genoma Espana Competition	7.7			7.7	7.8	15.5	*0.7*
C. difficile / H1N1	0.4			0.4	0.4	0.8	*0.0*
New Technology Development	9.6			9.6	9.7	19.3	*0.9*
Cancer Stem Cells Consortium	2.6	8.0	14.4	25.0	60.0	85.0	*4.1*
Advanced Technology Innovation Through Discovery	0.4	1.6		2.0	2.0	2.0	*0.1*
	587.1	**51.8**	**76.8**	**715.7**	**969.7**	**1,683.4**	*80.9*
S&T Innovation Centres	**102.4**	**12.0**	12.0	**126.4**	**47.1**	**173.5**	*8.3*
Genome Centres Operations	**57.9**	**5.5**	4.5	**67.9**	**68.6**	**136.5**	*6.6*
GENOME CANADA OPERATING EXPENDITURES	**70.7**	**8.0**	8.0	**86.7**		**86.7**	*4.2*
Total Disbursements	**818.1**	**77.3**	**101.3**	**996.7**	**1,085.4**	**2,080.1**	*100.0*
Excess (Deficiency) of Receipts over Disbursements	**34.3**	**(24.6)**	**(3.7)**	**6.0**			
Opening Cash Balance		**34.3**	**9.7**				
Closing Cash Balance	**34.3**	**9.7**	**6.0**	**6.0**			

Source: Genome Canada, Genome Canada Corporate plan 2011-2012, Ottawa, 2012.[5]

Figure 2.2 shows the inflow of funds from the federal government and the range of programs and projects funded over the first decade or so of the company's operations.

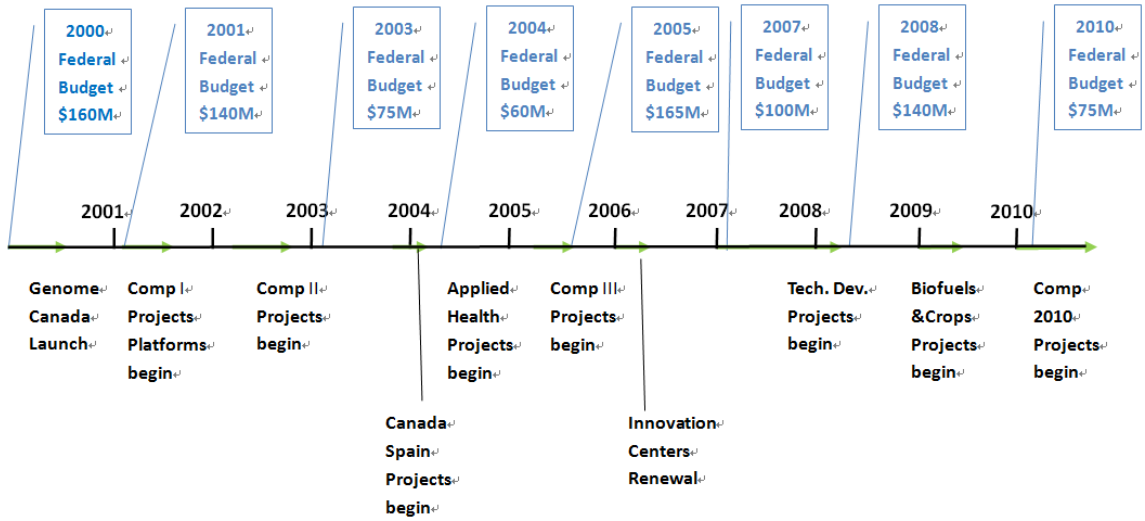


Figure 2.2 The flow of funds and investments

Source: Genome Canada, Genome Canada Corporate plan 2011-2012, Ottawa, 2012.[5]

From the data above and various financial reports, the overall efforts of GC can be summarized by the following:[6]

- $2 billion was invested, with more than half secured from partners;
- 156 large scale research projects across the life science sectors (see appendix 4 for the list of projects and their key operating dimensions);
- five world-class S&T Innovation Centres;
- more than 200 project leaders**,** who have developed the skills to manage complex science knowledge into application;
- more than 4,500 research publications; Canada ranks fifth in the world in terms of scientific impact, and fourth in the world in research related to science and society;
- more than 20 companies created;
- more than 10,000 highly skilled people employed; and

- more than 350 patent applicants and patent awards, and 24 license agreements; Canada ranked first in the multi-criteria ranking for intellectual property in genomics in 2005–2007.

2.4. **Selection Process**

As shown in table 2.4, Genome Canada has engaged in four large-scale, open research competitions, commonly named competitions I, II, III and the Applied Genomics in Bio-products and Crops (ABC) competition. The rest of the funding allocations were to directed projects/programs (called 'other' in this study) that were more directly managed and coordinated by Genome Canada or the centres.

Table 2.4 The large-scale open competitions

	Start Date	Total approved budgets	Number approved projects
Competition I	April 4, 2001	$136 million	17
Competition II	July 19, 2001	$155.5 million	33
Competition III	July, 2004	$346 million	33
ABC	April, 2008	$112 million	12
Total		$749.5 million	95

Source: Calculation from Genome Canada Corporate plan 2011-2012, Ottawa, 2012.[5]

Competition I was announced on April 4, 2001. An investment of $136 million was allocated in support of 17 large-scale research projects and five science and technology platforms across the country.[7]

On July 19, 2001, Competition II provided funding for several large-scale genomics research projects and their related science and technology platforms. A total budget of $155.5 million was made available to the 33 selected projects in April 2002. [8]

Genome Canada introduced Competition III in July 2004 and reported on August 25, 2005 that $346 million was invested in 33 large-scale projects for the duration of 3 to 4 years.[7]

Before the end of Competition III, Genome Canada engaged in a strategic research theme development exercise, involving a call for position papers from groups of scientists. In April 2008[7], GC announced a competition focused on applied genomics research in two related themes: 1) bio-products; 2) crops, hereafter called the Applied Bioproducts and Crops (ABC) competition. Projects in the bio-products theme were designed to "employ genomic and proteomic approaches to understand and manipulate the underlying biological processes exploited in the production of economically viable and environmentally sustainable bio-products. Three areas were targeted: feedstock optimization; microorganisms for sustainable processing technologies; and value added bio-products. Projects in the crops theme were required to foster an improved understanding of systems that govern plant growth, development and performance. Funded projects cultivated a comprehensive understanding of the genetic and physiological factors that contribute to the underlying biological processes of Canadian crops." Three areas were targeted: basic plant genomics; applications of plant genomics; and agriculture and food production sustainability. Results of this competition were announced on April 20, 2009: $112 million was invested in 12 new research projects.[9]

As discussed above, the selection process involves letters of intent which are vetted and approved for full application. This is then followed by submission of full proposals which are evaluated through peer review. The performance data in table 2.5 suggests the systems have operated somewhat differently in the different competitions. Competition I generated the most initial interest but as is common with a new grant program, many of the proposed ideas were not appropriate to the mandate of Genome Canada; in the end, the agency culled almost three-quarters of the ideas at the LOI stage. Other competitions only selectively culled at this stage. A second point of departure is in the submission of a proposal—many project leaders withdraw and do not submit a formal funding application due to the time and resource commitment of developing the full application. Once a project gets to peer review, its likelihood of receiving funding is quite high—ranging from 35% to 55% (and likely also if matching funding is easily arranged). Overall, 517 ideas were identified in LOIs (or registrations), leading to 213 proposals, 45% of which were accepted and funded, leading to an overall 18% conversion rate of ideas (at the LOI stage) into funded research.

Table 2.5 The flow of proposals in the open competitions

	Competition I	Competition II	Competition III	ABC Competition	Totals
Letters of Intent /Registrations	275	67	117	58	517
Full Proposals invited	73	64	93	48	278
Full proposals submitted for peer review	31	62	93	27	213
Approved projects	17	33	33	12	95
Success rates					
% of LOIs invited for full proposal	26.5%	95.5%	79.5%	82.8%	53.8%
% invited full proposals actually submitted	42.5%	96.9%	100.0%	56.3%	76.6%
% submitted proposals approved	54.8%	53.2%	35.5%	44.4%	44.6%
% of LOIs becoming approved projects	6.2%	49.3%	28.2%	20.7%	18.4%

Source: Phillips and Warren (nd) drawn from Genome Canada.

Due to the structure of Competition I, its emphasis was primarily on supporting large-scale projects. In fact, beyond the broad goals of the project proposals being large-scale, genome-wide, and in a sector considered important to Canada, there are no explicit references to project content at all.

Competition II provided a lot of details, guidelines and also began to place more of an emphasis on GE^{3}LS. Whereas the first competition simply asked each centre to have a program in place to deal with GE^{3}LS related issues, Competition II proposed that projects with a strictly GE^{3}LS focus as well as science projects with embedded GE3LS research could be submitted for funding.[5]

Competition III was marked with some significant changes in its preamble. GC announced it would accept applications from Genome Centers for large-scale research projects in genomics or proteomics for either three or four years in duration.[9] Genome Canada specified that proposals should be of such scale and scope that they cannot

currently be funded at internationally competitive levels through other existing mechanisms. Each project was now required to have one or more GE^{3}LS experts as a co-applicant, collaborator, or advisory committee member.[7] Also, an entire section in the preamble was dedicated to social and economic benefits of the research.

The ABC competition further developed the focus on GE^{3}LS by providing more detail about the format of the plan needed by project proposals to address GE^{3}LS issues. It directed that project proposals look at how GE^{3}LS work could enhance the research and realize maximum benefits. The guidelines asked applicants to integrate GE^{3}LS issues into the scientific components of their proposals, a concept absent from previous competitions. [10] The ABC competition guidelines became more precisely worded, exchanging words like "economic growth and social benefits" for "product and service development." [11]

2.5. **Past Evaluations of Genome Canada**

Genome Canada has been extensively reviewed. This section summarizes the nature and scope of the various reviews undertaken so far.

2.5.1. **KPMG Evaluation of Foundations**

This consultancy report, prepared for the Treasury Board Secretariat, presents the findings of an evaluation of the use of foundations (i.e. special operating enterprises) as instruments of public policy. This study was conducted by KPMG LLP on behalf of the Government of Canada between September 2006 and January 2007.

The study was triggered by the government's commitments to the Standing Senate Committee on National Finance and Standing Committee on Public Accounts to undertake an evaluation of the use of foundations as tools for the delivery of public policy, particularly with respect to the use of up-front conditional grant assistance. Genome Canada and five other foundations were the target of this review.

The collection of information for this evaluation relied upon four inter-related lines of enquiry, as shown in Figure 2.6

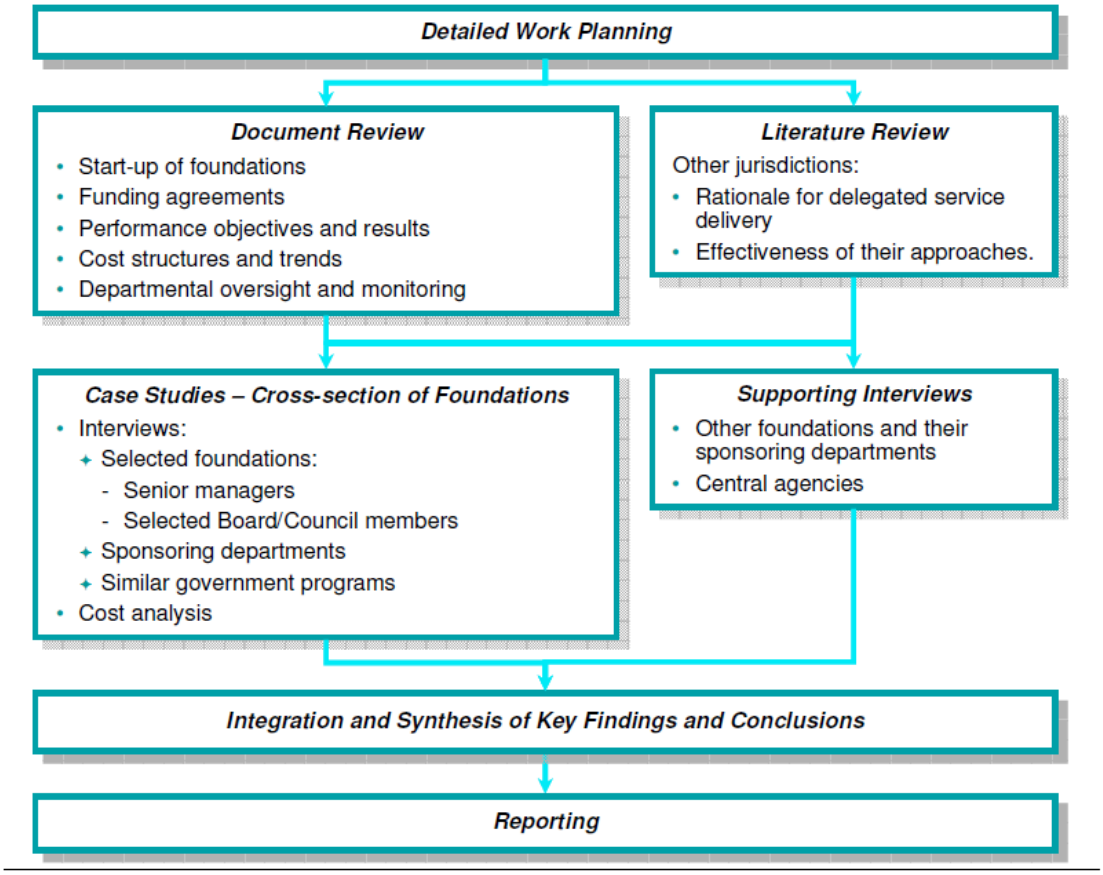


Figure 2.6 KPMG methodology for evaluating foundations

Source: KPMG LLP, KPMG Report: Evaluation of Foundations, Prepared for Treasury Board Secretariat. Ottawa, March 2007.[2]

This approach was designed to provide information from multiple sources to enable the evaluation issues to be assessed from several perspectives and to better understand the positions advanced by participants who are most closely involved with the use of foundations for public policy purposes.

The study had to be completed within a relatively short time period (Sept. 2006, to Jan. 2007), which necessitated a concentrated approach to data collection.

The evaluation team started with a review of the broad range of documentation on the government's use of foundations to achieve policy goals, the evolution of the terms and conditions under which foundation funding was been provided, and the results achieved by various foundations.

A series of six case studies of selected foundations were used as one of the two core data collection and analysis methods in the evaluation. The six case-study foundations were: Canada Foundation for Innovation (CFI); Canada Millennium Scholarship Foundation; Genome Canada; Aboriginal Healing Foundation (AHF); Green Municipal Fund (GMF); and the Pacific Salmon Endowment Fund Society. The case studies were used to obtain insights into the appropriateness, effectiveness and costs of specific foundations, which were used, in conjunction with findings from interviews with other foundations and stakeholders, to identify common characteristics, themes and conclusions applicable to most, or all.

KPMG reported on three aspects of the government's use of foundations. First, they examined the appropriateness of the foundation model as an instrument of public policy, concluding that the model exhibited generally strong degrees of alignment with the guiding principles published in Budget Plan 2003. [12] Second, they examined the effectiveness of the foundations, reporting on their progress against objectives, coordination with related government programs, alignment with government policy goals and their accountability mechanisms. The general conclusion was that they were doing well on all measures, albeit with some range of effectiveness. Third, they examined the operating and administration cost structures, focusing on structured and transparent processes for reviewing and selecting projects to support, and supporting systems for project tracking and financial management. The conclusion was that their operating and administration costs are driven by needs to efficiently manage project workloads and to provide timely support for governance and accountability requirements. Foundation resource levels and costs appear to be closely matched to, or follow, the trends in the project workloads.

In effect, KPMG offered an organization and operational review of the processes and structures, but did not undertake any specific analytical assessment of the fit of those processes to the overall goals.

2.5.2. **Risk Management Policy**

Genome Canada developed internally an integrated risk policy as a high level document outlining Genome Canada's approach and strategy towards Integrated Risk Management (IRM)[13]. Given that a Risk Management Policy must be able to 'stand the test of time' and be robust enough to withstand scrutiny from regulatory and/or legislative bodies, the Policy is broad in scope.

Risk management includes a risk methodology, risk profiles and related actions that will, by nature, change over time to reflect organizational changes and changes in risk profiles.

The Policy and related risk and action plans are applied to all operational aspects of the organization and considers external strategic risks arising from the external operating environment as well as other internal operational risks.

Although Genome Canada is not able to control external factors such as government priorities, they are considered and addressed as much as possible.

Methodology

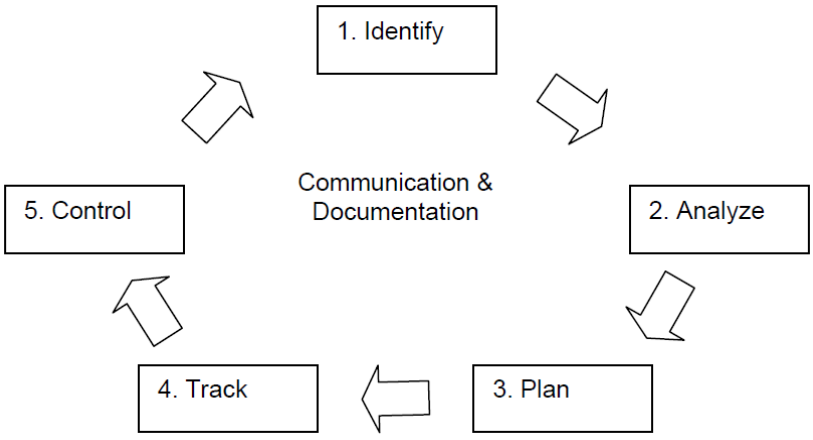


Figure 2.7 Five-step Genome Canada Risk Management Framework
Source: Risk Management Policy[13]

This internal policy entails an ongoing series of operational evaluations used to manage and safeguard the entity. While critical to effective operations, the Policy involves more tactical evaluation than strategic review.

2.5.3. **KPMG Report on Genome Canada**

In 2008 KPMG was contracted by Genome Canada to do an overall evaluation of the impact of Genome Canada investments. GC is directed to undertake an evaluation every five years as a requirement of their funding agreement with industry Canada. Table 2.8 provides a breakdown of the large-scale projects and S&T platforms that had been funded as of June 3, 2008, broken down by region and sector of application.

Table 2.8 Funding Allocation

Sector	BC		Alberta		Prarie		Ontario		Quebec		Atlantic		Canada Total	
Agriculture	2	$9,164	1	$6,806	3	$17,551	1	$814			1	$1,925	8	$36,266
Environment	1	$2,305					2	$8,416	1	$3,756	1	$2,083	5	$16,564
Fisheries	2	$10,553									2	$10,950	4	$21,505
Forestry	2	$15,429					1	$2,327	2	$11,385	1	$910	6	$30,055
GE3LS	2	$1,630	1	$1,330	1	$1,663	3	$9,674	2	$2,430			9	$16,734
Health	18	$80,377	2	$9,175	2	$21,866	22	$175,285	22	$126,368	2	$6,857	68	$419,978
New Technology Development			1	$2,283	1	$8,564	3	$11,745					5	$22,597
S&T Platforms	4	$26,545	1	$5,680	1	$5,024	2	$17,763	1	$23,801	1	$5,805	10	$84,624
Total	31	$146,003	6	$25,274	8	$54,668	34	$226,024	28	$167,740	8	$28,530	115	$648,323

Source: KPMG Report[14]

This evaluation focused on the impact of the funding allocations. The methodology involved a review of internal documentation and databases, web-based surveys and interviews and a partial cost-benefit analysis of GC research investments and outcomes. As outcomes based approach, the analysis did not directly assess the fit between the research funding decisions and the strategic goals of GC.

2.5.4. **Performance, Audit and Evaluation Strategy of Genome Canada (PAES)**

In 2008, Genome Canada articulated a full performance, audit and evaluation strategy (PAES); while this updated in 2013, we focus on the earlier version here as it was the one

operating during our study period. Figure 2.9 illustrates the elements.[15] The strategy was developed as a high level framework which addresses key elements that Genome Canada had implemented or planned to put in place to ensure accountability in the achievement of objectives from the perspective of performance, audit, evaluation and reporting.

PERFORMANCE REPORTING			
Corporate Plan	***Annual Report***	***Special Publications***	***Web Site***
PERFORMANCE MONITORING	**AUDIT**	**EVALUATION**	
Project Selection	***Financial Audits***	***Results Based Management And Accountability Framework (RMAF)***	
Project Monitoring	***Recipient Audits***	***Logic Model***	
Interim Review	***Compliance Audits***	***Evaluations Every Five Years***	
Performance Indicators	***Performance Audits***		
Final Project Reports			
RISK MANAGEMENT			
Risk Management Framework	***Operations and Management***	***Governance Regime***	

Figure 2.9 PAES

Source: Genome Canada - Performance, Audit and Evaluation Strategy, 2007.[15]

These processes are designed to contribute to more effective operations and to ensure compliance to the funding agreements signed with Industry Canada with respect to the use and accounting of funds received from the federal government. Genome Canada also signs individual funding agreements with each of the six Genome Centers where the undertakings agreed to with Industry Canada are essentially replicated.

The PAES is comprised of three key frameworks: 1) Performance monitoring and measurement; 2) Audit; and 3) Evaluation. All elements provide a foundation for strengthening internal management.

2.6. **Implications for this research**

As reviewed above, while there have been some efforts to assess the operations of the organization related to its goals and objectives, this work has been mostly in the form of institutional audits and qualitative assessments. This study extends that work. It offers an empirical, quantitative assessment of the fit between the institutional goals and objectives and the funding allocations of the organization to determine the relative balance and impact of the diverse objectives on their core activity of funding research.

3. Evaluation in the Policy Literature

3.1. **Definition**

Evaluation is a critical part of the public policy system, as it helps to define problems, delimit options, aid with decision making and improve operational efficiency. Evaluation is defined as the systematic determination of merit or worth using criteria against a set of standards.[16]

At the individual level, evaluation can be the formal determination of an individual's job-related actions and their outcomes within a particular position or setting. In financial trading, its objective is to assess the extent to which an individual added wealth to a firm and/or its clients, and whether his or her achievement was above or below the market or industry norms, also called measurement.[17]

At the organizational level, evaluation is a critical link in Simon's (1997) ends-means causal chains. Only with organizations, the focus is on how specific activities or processes contribute to the goals of the institution or agency.

The design of a particular evaluation approach depends on the actors involved and the situation.[15] Standards and principles of evaluation give some sense of direction and the base of ethical norms, commitment and integrity. In our study, the stated goals of GC are the foundation of the whole process for project evaluation.

In the Government of Canada, evaluation is the systematic collection and analysis of evidence on the outcomes of programs to make judgments about their relevance, performance and alternative ways to deliver them or to achieve the same results.

Evaluation provides Canadians, Parliamentarians, Ministers, central agencies and organizational heads an evidence-based, neutral assessment of the value for money (i.e. relevance and performance) of federal government programs.

Evaluation:

a. supports accountability to Parliament and Canadians by helping the government to credibly report on the results achieved with resources invested in programs;

b. informs government decisions on resource allocation and reallocation by:

    i. supporting strategic reviews of existing program spending, to help Ministers understand the ongoing relevance and performance of existing programs;

    ii. providing objective information to help Ministers understand how new spending proposals fit with existing programs, identify synergies and avoid wasteful duplication;

c. supports deputy heads in managing for results by informing them about whether their programs are producing the outcomes that they were designed to produce, at an affordable cost; and,

d. supports policy and program improvements by helping to identify lessons learned and best practices.[18]

Evaluation products means any output of the departmental evaluation function, which may include, but is not limited to, the departmental evaluation plan, terms of reference for individual evaluations, evaluation assessments, evaluation frameworks, evaluation reports, and advice.[18]

3.2. **Literature Review**

In an early paper on performance evaluation, Arvidsson (1986) focused on the pressures facing public services. He asserted that government performance evaluation could be measured in several ways, either by examining objectives, timing and the procedures of international administration. [19]

King (1987) asserts that research evaluation "makes use of a variety of indicators to draw as complete a picture as possible of the complex aspects that account for the performance of research".[20]

Peter Henry Rossi (2004) defined program evaluation as the use of social research procedures to systematically investigate the effectiveness of social intervention programs, adapted to the political and organizational environments and designed to inform social action in ways that improve social conditions. Comprehensive evaluation is an assessment of a social program that covers the need for the program, its design, implementation, impact, and efficiency. [21]

The differences between policy analysis and policy evaluation are widely known but increasingly unrecognized. Geva (1999) compares policy evaluation and policy analysis in terms of concept, methodology, problems and data description. Evaluation tends to adopt a focus on the analyst/process which is being used to make policy choices.[22]

Theory-based evaluation (TBE) has become widely discussed and occasionally practiced in the recent years. Birckmayer (2000) [23] identified evaluations may be needed beyond the regular operational assessment. Supporters think this approach will help to explain how and why formal project assessments predict the results. Very often, this type of evaluation will follow each step in a sequence to see whether the expected steps actually occurred.

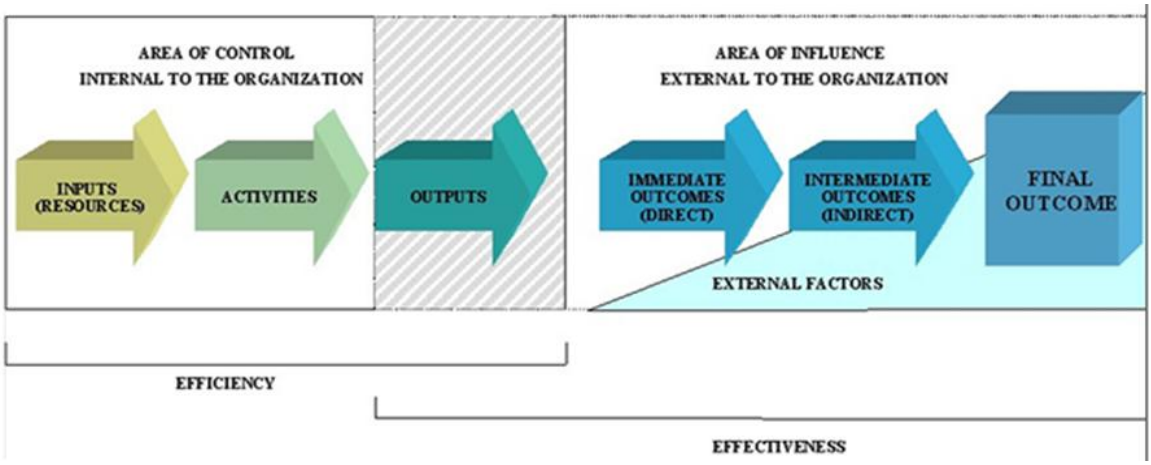


Figure 3.1 Treasury Board of Canada Outcomes Management Framework

One way to look at the challenge of evaluating research systems is through an outcomes management framework, as shown in Figure 3.1. In this context, it is possible to see that evaluation could focus on efficiency and effectiveness, with efficiency analysis investigating the causal path between inputs, activities and direct outputs. With respect to Genome Canada, the inputs could be viewed as the allocation of funds from Industry Canada, as illustrated in Figure 2.3. The focus of this work is on the 'activities' undertaken by Genome Canada to allocate funds to specific science projects (the outputs).

Luukkonen (2002) notes that research evaluation also is connected with the assessment of performance of applicants and on the embedded decision-making sub-systems, such as peer review.[24]

Michael Quinn Patton (2002) asserts "a successful evaluation emerges from the special characteristics and conditions for a particular situation—a mixture of people, politics, history, context, sources, constraints, values, needs, interests, and chance. Despite the rather obvious, it is not at all obvious to most stakeholders who worry a great deal about whether an evaluation is being done right. Indeed, one common objection stakeholders make to getting actively involved in designing an evaluation is that they lack the knowledge to do it right."[25]

3.3. **Overview of Evaluation Methods**

In essence, performance evaluation is described as comparing results against objectives, which will vary with different situations. It could also be applied in many ways. Here is a list of various evaluation methods (table 3.3).

Table 3.1 Characteristics of Evaluation Methods

Methods	**Description**
Peer review/expert judgment	Qualitative review, opinion, and advice from experts on the subject being evaluated based on objective criteria
Case study	Information through a narrative about the subject
Historical tracing	A series of interrelated events either going forward from the research of interest to downstream outcomes or working backward from an outcome along a path that is expected to lead to precursor research
Network analysis	Visual mapping and measurement of relationships and linkages among researchers, groups of researchers, laboratories, or other organizations
Benchmarking	The systematic comparison of practice, status, quality, or other characteristics of programs, institutions, regions, countries, or other entities using a selected set of performance measures
Survey	Obtaining information directly from people about their ideas, opinions, attitudes, beliefs, preferences, concerns, plans, experiences, observations, and virtually any other issue; interviews, document review, literature review
Technology commercialization tracking	The new energy-efficient technologies developed through R&D projects sponsored by the program, which may include research cost-shared with an industry
Benefit-cost case study	Applied research and technology programs with well-defined goals that lend themselves to at least partial economic interpretation and analysis, though assessed benefits and costs often extend beyond economic effects
Econometric methods	A variety of statistical and mathematical tools and theoretical models to analyze and measure the strength of functional relationships that underpin a program and to analyze and measure a program's effects on firms, industries, innovation, and the economy

Source: USA Department of Energy (2007).[26]

To date, Genome Canada has used a range of these methods. The most prominent choices have been document review, peer review (used for competition I, II, III, ABC) and case study. The KPMG Evaluation of Foundations evaluation team reviewed a broad range of documentation on the government's use of foundations to achieve policy goals, the evolution of the terms and conditions under which foundation funding has been provided, and the results achieved by various foundations. They also undertook case studies to obtain insights into the appropriateness, effectiveness and costs of specific foundations.

The internal processes detailed in Chapter 2, Figures 2.2 illustrate the role of historical tracking in assessing the system. KPMG's review in 2009 used a mixed method approach, including peer-reviewers, expert judgment, survey and benefit cost.

Our study applies some of the insights from the econometric approach exemplified by Lusk to empirically evaluate the fit between goals an allocation decisions in the context of the open competition and internal project development processes. By testing several factors, we intend to evaluate the relationship between the chosen projects and stated goals of each funding initiative. To date, econometric methods have not been applied directly to the Genome Canada investments. In other areas, these tools have been widely used to identify the causal links between inputs and outputs. Lusk et al (2005), for example, used a meta-analysis to evaluate the impact of consumer willingness-to-pay (WTP) or willingness-to-accept (WTA) values for various novel food products. The goal is to generate a set of findings about consumer WTP/WTA for food that are based on the results of a single study, but to provide policy makers with a nuanced summary of a body of work. "For example, a dummy variable was created to identify whether the valuation was from a study that strictly elicited WTA. Finally, several variables were created to describe the good valued in each of the studies, including the food type and whether the food provided any direct benefit, such as enhanced nutrition, to the consumer". [27]

Narongrit (2010) used grouping method and pilot 3D location as an evaluation method to assess academic ranking as a means of allocating resources. The Office of the Higher Education Commission (OHEC), Ministry of Education in Thailand had considered the university rankings to be measured among the academic community, in the purposes of assigning budget allocations for academic promotions. [28]

4. Methodology, Model & Data

4.1. The Logic of the Model

This chapter lays out the logic for assessing the operational fit between Genome Canada's investment program between 2001 and 2012 and the organizational mandate. The basis assumption is that we will find a positive and significant fit between the goals and the nature of the funding allocations.

The goal is to undertake a strategic analysis. In order to model the process appropriately, we have laid out the Genome Canada process logic.

Figure 4.1 Logic of the Process

The funding agreement between the Government of Canada and Genome Canada lays out the organizations objectives. Those objectives are taken as high level criteria by which the organization will allocate the funds provided to GC.

The government's overall science and technology policy goal is the production of scientific knowledge and the advancement and commercialization of technical knowledge.

From 2000 to 2012, the specific objectives of Genome Canada are: (1) developing and implementing a coordinated strategy; (2) bringing together industry, governments,

universities, research hospitals and the public to support large-scale genomics and proteomics research projects; (3) providing accessibility to Science & Technology Platforms to researchers; (4) assisting in attracting co-funding for projects from both domestic and international investors; and (5) sustaining leadership. [2]

Those goals then translate into five core objectives that should be reflected in the funding allocation decisions:[29]

Objective 1 is to develop and implement a coordinated genomics research strategy. In practical terms, this translated into a series of internal processes to assess and identify coordinated strategies for genomics research to enable Canada to become a world leader in areas such as sector health, agriculture, environment, forestry, fisheries, tech and GE3LS.

Objective 2 is about providing leading-edge technology. Operationally, this involves the provision of leading-edge technology to researchers in all genomics-related fields.

Objective 3 is to support large-scale research. In effect this is a scale issue. Given the nature of the Genome Canada database we have generated (i.e. not including the projects that were rejected), we cannot show this effect inside our data. One way to see scale is to compare the allocations by Genome Canada with allocations on genomics-related research by the Canadian Institutes for Health Research (CIHR) and the Natural Science and Engineering Research Council (NSERC). Data in table 4.1 shows the average size of GC allocations are about 10 times the size of the average CIHR grant and about 65 times larger than comparable awards by NSERC.

Table 4.1 Grants for genomics related research, 1999-2012

	Granting Period	# projects funded	Total value of direct outlays	Average $000/project
CIHR	1999-2014	1370	$572.1 M	$ 417.6
NSERC	1999-2012	1130	$ 75.2 M	$ 66.5
GC	2001-2012	156	$ 682.6 M	$4,375.5

Source: Author's calculations using data derived from the CIHR and NSERC Funding Decision Databases, Aug. 6th, 2013 Objective 4 is to assume GE3LS leadership and to

communicate more effectively with Canadians. The assumption of leadership in the area of ethical, environmental, economic, legal, social (GE3LS) and other issues related to genomics research and the communication of the relative risk, rewards and successes of genomics to the Canadian public can be assessed by the role and position of GE3LS in the structure of each competition and in the related projects.

Objective 5 is to encourage investment by others. In practice, this can and should be measured by whether the projects leverage co-funding from non-governmental sources, including international sources.

The purpose is to explore the influence of key factors in the selection and allocation of funds to projects.

While we are ultimately concerned about efficacy and accountability of the choice systems used by Genome Canada, the key processes are not directly measureable—they are effectively in a black box. Nevertheless, they are the indirectly discernible through examining the information available at the time of the decision making and the resulting allocations of funds.

An econometric approach was used to fit proxies for the stated objectives to the share of the portfolio allocated to each project. .

A series of regressions will be employed to determine the proportion of the funding allocations that are explained by the objectives. The residual could be interpreted as the influence of soft factors, like the personal preference of the reviewers and Genome Canada staff, the cognitive bias of the various decision makers, the context of the specific science platform and the uncertain environment.

4.2. **Data Sources**

As shown in table 2.4 in chapter 2, Genome Canada has engaged in four large-scale, open research competitions, commonly named competitions I, II, III and the Applied Genomics in Bio-products and Crops (ABC) competition, and the other competitions.

The funding data is mentioned in chapter 2. As of 2012 Genome Canada had committed $915 million in funding and researchers had secured approximately an additional $1,085 million in co-funding, representing a total investment of over $2 billion in completed or planned genomics research in Canada. We have used that data, allocated by specific project, to calculate project shares of funding and used this as the dependent variable.

The independent variables that are assessed for fit with funding decisions are discussed below.

4.3. **Basic Equation**

The basic equations in the model involve running regressions with the allocation decisions as the dependent variable and the key organizational and program objectives as the independent variables.

The basic estimation equation is:

$Y = a + b_1 * GE^3LS + b_2 *$ Technology $+ b_3 *$ International co-funding $+ b_4 *$ PI reputation $+ b_5 *$ Institution research intensity $+ b_x *$ competition, section and regional dummies

(4.3.1)

The following variables have been chosen to describe the potential relationship between the different variables.

4.4. **Dependent Variables**

Two dependent variables have been tested, that is Y_1 (GC-total) and Y_2 (Open-com). The regression using the total pool of investments provides insights into the performance of the organization across the portfolio of investments. This portfolio is chosen through two discrete systems. The main portion of the funding is allocated through open competitions, where investigator-led teams submit competitive proposals that are adjudicated through a competitive peer-review process. The rest of the portfolio involves directed projects, where Genome Canada, one of the regional centres or a partner has developed a project to fit a specific strategic or tactical need. These projects are internationally peer-reviewed but there is little in the way of competition in the process. The second regression tests to

see how the choices in the open competitions conform to the stated goals of Genome Canada. By reduction, any difference in fit between the open competitive process and the overall pool would tell us something about the efficacy of the process of developing directed projects

Table 4.2 Explanations of dependent variables

GC-goals Objective	Subject	Unit	Description	Calculation	Source
Allocation of Fund	Y_1: GC-total	%	% share of GC contribution of each project in the total fund pool of all Genome Canada contribution	$Ai/\sum_{i=1}^{n} Ai$ (%) (i~[1,156], n=156)	Genome Canada Reports[1]
	Y_2: Open-com	%	% share of GC contribution of each project in the open pool of GC contributions in I, II, III and ABC competitions.	$Ai/\sum_{i=1}^{m} Ai$(%) (i~[1,95], m=95)	

The percentage share of each project in the total fund pool is a way to measure the allocation of funding. That is for each project, the assigned fund will share Yr% of the funding pool in both the total and open competitions.

The subject Y_1 GC-total is the percentage share of GC contribution of each project in the total fund pool of all Genome Canada contributions. This pool involves 156 projects which shared $683 million funds invested by Genome Canada. It is calculated as the GC contribution dollar of each project (Ai) as a percent of entire portfolio. While Genome Canada has invested $996 million, about one third of the commitments and disbursals is for infrastructure and operations and not to fund research projects.

The subject Y_2 open-com is the percentage share of GC contribution of each project in the open pool of GC contributions in Competitions I, II, III and ABC. From the calculation, we could know that the total open pool $\sum_{i=1}^{m} Ai$ (i~[1,95], m=95) equals $485 million. Ai is the GC contribution of each project.

4.5. **Core Independent Variables**

Five core variables have been identified as conforming to four of the objectives:

Table 4.3 Explanations of Behavior Independent Variables

GC-goals Objective	Subject	Unit	Description	Calculation	Source
(a) Sustain leadership and coordinated strategy	X_1 PI (lead Harzing index)	Index	Principal Investigator(PI) research capability: measured by HI index (collected by 2012.7)	Lead Harzing Index-HI Index	www.harzing.com
	X_2 Research intensity	dollar	Host institution research capability: measured by Total Research Dollars (10000$ per full-time faculty member)	Total Research Dollars (10,000$ per full-time faculty member)	Appendix II
(b) Support GE3LS	X_3 GE3LS		Whether the project supports GE3LS	Yes=1; No=0	Genome Canada Reports [1]
(c) Encourage PPP (public-private partnership)	X_4 International co-founding		International co-funding source	Yes=1; No=0	
(d) Provide leading-edge technology	X_5 Technology		Whether the project is in a technology development activity and represents the leading-edge	Yes=1; No=0	

4.5.1. **PI and Research Intensity as a measure of Leadership**

The coordinated genomics research strategy is designed to support leadership, which is assumed for this analysis to be represented by the Principal Investigator's (PI) research capability measured by the Harzing Index (HI) index (X_1) and a variable that measures the research intensity of the host institution (as measured using the MacLean's research funding measures) (X_2).

The HI index (X_1) was proposed by J.E. Hirsch[30] in a paper entitled "An index to quantify an individual's scientific research output".[1] It is defined as follows: "A scientist has index h if h of his/her N_p papers have at least h citations each, and the other (N_p-h) papers have no more than h citations each." It aims to measure the cumulative impact of a researcher's output by looking at the amount of citations among the most highly cited parts of his/her work. The calculation tool Publish or Perish[2] calculates and displays the *h* index proper, its associated proportionality constant *a* (from $N_{c,tot} = ah^2$), and the rate parameter *m* (from $h \sim mn$, where *n* is the number of years since the first publication).

One option to see the trend is through a scatter-plot. Using the scatter-plot procedure, (by typing *scatter yvar xvar*, || lfit *yvar xvar)* we generated a scatter-plot with PI along Y_1 GC-total and Y_2 open-com.

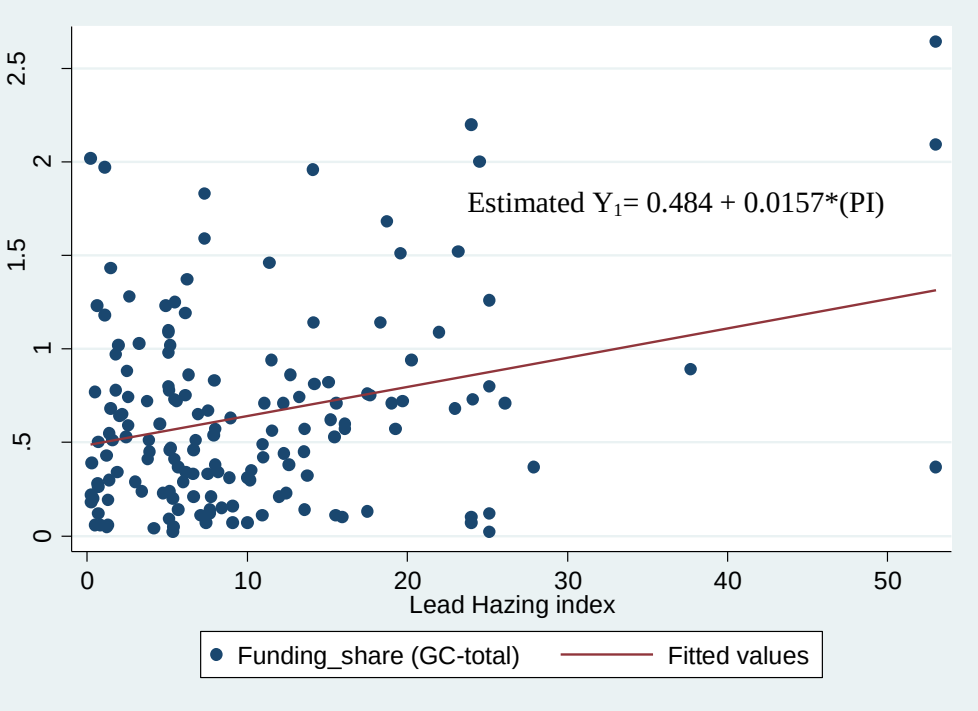


Figure 4.2 Y_1-GC-Total along PI

[1] arXiv:physics/0508025

[2] The properties of the h-index have been analyzed in various papers; see for example Leo Egghe and Ronald Rousseau: An informetric model for the Hirsch-index, Scientometrics, Vol. 69, No. 1 (2006), pp. 121-129.

From Figure 4.2, the assumption of the positive relationship between PI and Y can be seen on the fitted values. The regression results are shown below.

Source	SS	df	MS		Number of obs =	156
					F(1, 154) =	14.51
Model	3.48582474	1	3.48582474		Prob > F =	0.0002
Residual	36.9861931	154	.240170085		R-squared =	0.0861
					Adj R-squared =	0.0802
Total	40.4720178	155	.261109792		Root MSE =	.49007

GC_total	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.015665	.0041118	3.81	0.000	.0075421	.0237879
_cons	.4839159	.0569839	8.49	0.000	.3713448	.5964869

Figure 4.3 Regress GC-total PI

Estimated Y_1= 0.484 + 0.0157*(PI) **(4.5.1.1)**
(8.49)* (3.81)***

This equation tells us that, all other things being equal, for every 1 unit increase in PI HI index, Y_1 is expected to increase by 1.57%.

Limiting the analysis to the open competitions, we find the slope and the intercept are statistically significant at 98% and 99% confidence interval respectively. For every 1 unit increase in HI, Y_2 is expected to increase 2.74%. Using the adj-R^2 we can see that about 7% of the variance in Y_2 is explained by the PI HI indicates.

Another way is to look at the individual variable character through descriptive statistics. Appendix I present the results of a histogram and codebook analysis. The codebook and histogram shows the "feel" of the PI. In this case, the PI HIs range from 0.2 to 53. The mean is not near the centre of the range; it is located at the end of first quarter of the range. Almost 90% of the index numbers were in the bottom half of the range. The distribution is not equal, which means it is not normal distribution.

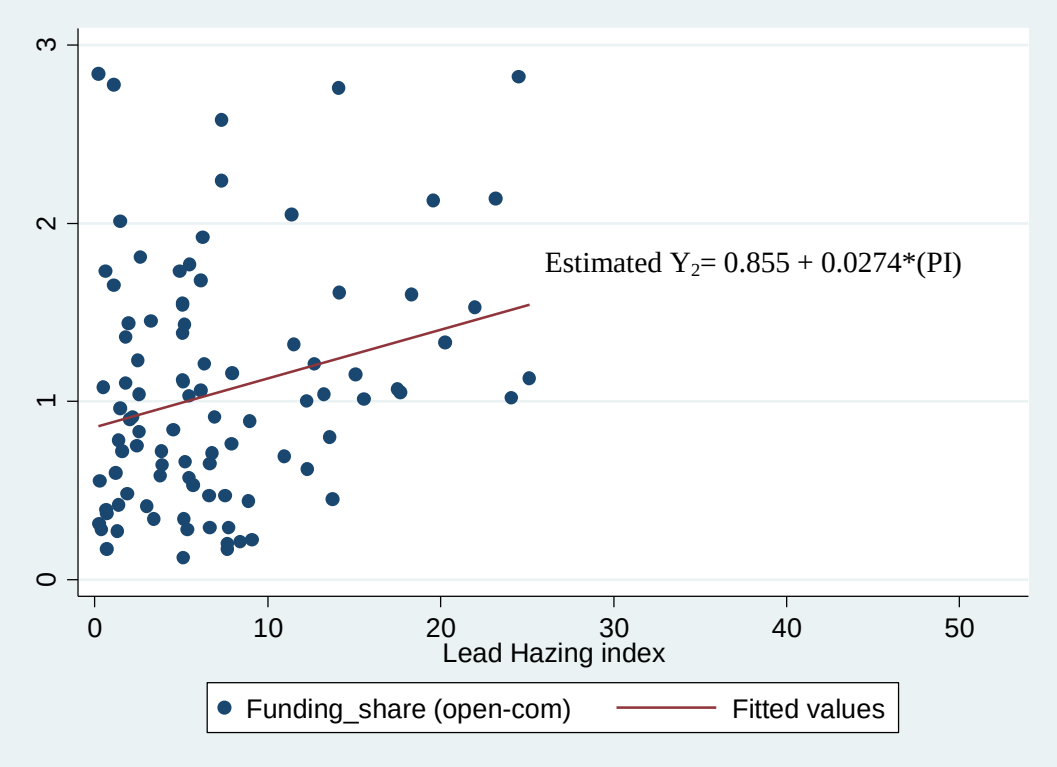


Figure 4.4 Y_2 Open-com

```
. reg  open_com PI
```

Source	SS	df	MS
Model	2.76180077	1	2.76180077
Residual	38.6088344	93	.415148757
Total	41.3706352	94	.44011314

Number of obs	95
F(1, 93)	6.65
Prob > F	0.0115
R-squared	0.0668
Adj R-squared	0.0567
Root MSE	.64432

open_com	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.0273926	.0106204	2.58	0.011	.0063027	.0484826
_cons	.8554103	.1009547	8.47	0.000	.6549343	1.055886

Figure 4.5 Regress Open-com PI

Estimated Y_2= 0.855 + 0.0274*(PI) **(4.5.1.2)**
(8.47)* (2.58)****

A second factor is institutional leadership. Given that one of the stated objectives of Genome Canada is to generate globally competitive research capacity, it would be

appropriate to assess whether prior institutional capacity is influential in determining the allocation of funds. The relative research intensiveness of the host institutions, as measured through the MacLean's institution research reports (X_3) is one way to rank the host institution research capability (see Appendix II).

The annual *Maclean's*[3] rankings assess Canadian universities on a range of performance indicators in six areas. We chose the Total Research Dollars reported in *Maclean's* (including income from sponsored research such as grants and contracts, federal, provincial and foreign government funding, and funding from non-governmental organizations) adjusted for the relative size of each institution (i.e. using a capitation formula based on full-time faculty).

The indicator Resources-Total Research Dollars is chosen to evaluate the research capability of the host institution, which is then rebased to 10,000 dollars per full-time faculty member. From the codebook in appendix II, the range of this variable is 0.43 to 3.51, with a mean of 2.51.

4.5.2. **GE3LS**

Objective 4 asserts GC seeks to generate leadership in the area of ethical, environmental, economic, legal, social (GE3LS) and other issues related to genomics research and the communication of the relative risk, rewards and successes of genomics to the Canadian public (X_4). Projects can either embody integrated research (INTERGE3LS) or can be stand alone. This is a dummy variable with a value of 1 if GE3LS is embodied in some way in the project and zero otherwise. Of the 156 projects, 11 are stand-alone GE3LS projects and 50 are INTERGE3L.

4.5.3. **Leveraged co-funding as Public-Private-Partnerships (PPP)**

Genome Canada established ambitious co-funding goals for their projects. The minimum threshold was 100% matching, in cash or in kind. All approved projects by definition met that goal. Over the past decade, GC has attracted $1 billion in co-funding to complement

[3] http://tools.macleans.ca/ranking2008/selectindicators.aspx

the $980 million committed by the Government of Canada.[4] There is little difference in leveraging among projects.

We were particularly interested whether public-private partnerships (PPPs) were influential. To test that, X_5 was defined as the presence of investment by an international co-founder. Projects with identified international partnership were coded one; projects with only domestic funding were coded zero.

4.5.4. **Technology**

The variable Technology (X_6) corresponds to objective 2, providing leading-edge technology. In the final report of Genome Canada, it tests whether the project is deemed to be in the "technology category" or not.[1] It is determined by the category factor, which could shown in the GC Database (in appendix VI). Operationally, X_6 involves the provision of leading-edge technology to researchers in all genomics-related fields through regional Genome Centers across Canada, which represents objective 2.

4.5.5. **Regional, sectoral and competition dummies**

Given that there were four competitions and the directed investments, seven priority research areas and six geographic regions, it is possible that context may have been a determining factor in the funding allocations. Table 4.4 shows how those factors have been converted into dummies to control for these technical factors. The only significant change we made was to combine Genome Alberta and Genome Prairie, on the basis that their activities were highly correlated; Genome Prairie, located in Edmonton, served the three Prairie Provinces until 2005, when Genome Alberta became an independent centre and Genome Prairie moved operations to Saskatoon. Since then they have collaborated closely on development and management of a range of successful projects, making it problematic to include them as fully independent contextual variables.

[4] 2012 Annual Report of GC

Table 4.4 Description of Dummies

Part	Variable = 1	Description	Calculation (Freq.)[5]	mean	% of fund
Sector	Health		82	0.52	62.26%
	Agriculture		16	0.096	8.55%
	Environment	Environment, energy, fishery	19	0.09	15.53%
	Forestry		11	0.071	6.21%
	Technology	Providing leading-edge technology	18	0.115	4.14%
	GE3LS	The research on the ethical, environmental, economic, legal and social (GE3LS) aspects of genomics.	11	0.071	3.32%
Region	BC		40	0.256	22.64%
	Prairie	Alberta, Saskatchewan and Manitoba	21	0.134	15.61%
	ON		52	0.339	37.06%
	Quebec		34	0.218	21.15%
	Atlantic		8	0.051	3.59%
Competition category	com1	Competition I	17	0.109	11.82%
	com2	Competition II	33	0.212	21.43%
	com3	Competition III	33	0.212	29..99%
	ABC	Applied genomics research in Bio-products or Crops(ABC)	12	0.077	7.77%
	Directed competitions	Other categories	61	0.391	28.99%
Total			156		

Source: Appendixes codebook and sum

Note: Tab X- STATA command in having Frequency and Percentage

[5] Codebook-STATA, Appendixes III

Sum X- STATA command in having mean

% $ of the fund is calculated by the original data in excel

From the above Table 4.4, the sum of dummies for each category above equals to one, as all variables cover all the possibilities in each category. For example, a project by definition must be in one of the regions (British Columbia, the Prairies, Ontario, Quebec and the Atlantic), sectors (health, agriculture, environment, forestry, tech, GE3LS) and Competition category (com1, com2, com3, ABC, Directed). To avoid over definition of the regression, at least one variable from each category is excluded in each regression. In the end, the extra detail offered by the six regions, seven sectors and five competition categories did not add much descriptive power. While all of the dummies are presented here and in tables 5.1 and 5.2, the regression results presented in tables 5.5 and 5.6 only involve the single largest variable in each category (i.e. Ontario, health and directed projects).Before exploring the relationship between funding share and project character, the whole data set was built using the above rules. The dataset of 156 projects is called data-full and is included in appendix VI.

Once the dataset was constructed, STATA (version IC/11.1) statistical package was used to estimate regressions.

The first step, even before running any multivariate regression, was to look at the individual variables and their distributions. To do that we looked at the histograms kernel density curves as presented in the appendixes. These show that most of these variables are not normally distributed.

Therefore we ran the ladder test for individual variables using the chi-squared test to identify the closest normally distributed transformation (Appendix IV). One of the ways to correct for this is to transform one of the variables. We chose to test whether a transformation would help. We transformed the PI variable into a log form (i.e. generated LGPI= log (PI)). The log transformation is a monotonic transformation, which keeps the order of the numbers, while transforming the distribution of the observations.

From the histogram of the two variables PI and log PI (Figure 13), we can see that the distribution of log PI is closer to normal distribution, compared to the skewed distribution of PI. Looking at the summary of both variables, we can see the standard deviations are much smaller for log PI. Using this information, we transformed PI and did regressions on each of them (see Appendix IV Figure 5).[6] Comparing the regression results from non-transformed variables with transformed variables, we found that the overall model fit (R^2) deteriorated 12% to 2.6%, so that the above transformed was not used in the formal regressions that follow.

The OLS method is chosen to estimate the model for two reasons. First, the lack of any obvious correlations between the independent variables suggests that the variables may be independently considered in the decision system. Furthermore, there was no obvious direction or effort to differentially assess and apply the independent variables in the decision system—i.e. Genome Canada does not direct specific weights be used nor does it provide any architectural design to the consideration of these variables. All variables are considered equally in the decision system, with weights being revealed through choice rather than assigned a priori. Thus, in absence of any other evidence to the contrary, the OLS was chosen as the most appropriate method of calculating the influence of these variables on the overall decisions.

[6] Appendixes Figure 14 Comparison of regression on PI and Log PI

5. Empirical Analysis & Regression

This section presents and discusses the multivariate model results.

5.1. Correlation Test

There were a number of issues that came up during the multivariate model building phase. Before testing the relationship between the Xi and Y, we tested to determine whether the independent variables were correlated and involve the risk of multicolinearity. The correlation matrix in Figure 5.1 is the test.

```
. correlate  PI research international interGE3LS health agriculture environment forestry fisheries tech GE3LS BC
> Prairie ON Quebec Atlantic com1 com2 com3 ABC directed_com
(obs=155)
```

	PI	research	intern~l	interG~S	health	agricu~e	enviro~t	forestry	fisher~s	tech	GE3LS
PI	1.0000										
research	0.1820	1.0000									
internatio~l	0.1935	0.1319	1.0000								
interGE3LS	-0.0215	-0.1013	-0.1764	1.0000							
health	0.1338	0.1825	0.1320	-0.1395	1.0000						
agriculture	-0.0944	-0.1353	-0.1171	0.0490	-0.3469	1.0000					
environment	0.0578	-0.2165	-0.0812	0.0305	-0.3596	-0.1111	1.0000				
forestry	-0.0780	-0.1039	-0.0556	0.1647	-0.2783	-0.0860	-0.0891	1.0000			
fisheries	-0.0863	-0.2101	0.0112	-0.0173	-0.1489	-0.0460	-0.0477	-0.0369	1.0000		
tech	0.0262	0.1416	0.0290	-0.2096	-0.3842	-0.1186	-0.1230	-0.0952	-0.0509	1.0000	
GE3LS	-0.0866	0.0930	-0.0145	0.3431	-0.2426	-0.0905	-0.0938	-0.0726	-0.0388	-0.1002	1.0000
BC	-0.1086	-0.1769	0.0687	0.1285	-0.0343	-0.0434	-0.0063	0.2052	0.0242	-0.0757	0.0093
Prairie	-0.1013	-0.1879	-0.0921	0.1441	-0.2685	0.3805	0.0516	-0.1040	-0.0556	0.0330	0.1108
ON	0.2368	0.3905	0.0974	-0.1631	0.1079	-0.1439	-0.0211	-0.1339	-0.1013	0.1632	-0.0403
Quebec	-0.0591	-0.0072	-0.0713	-0.0760	0.1566	-0.1208	-0.0261	0.0512	-0.0745	-0.0948	-0.0251
Atlantic	-0.0274	-0.1952	-0.0733	0.0156	-0.0438	0.0339	0.0283	-0.0571	0.4205	-0.0788	-0.0601
com1	-0.1534	-0.0442	-0.0020	-0.1560	0.0003	-0.0451	0.0166	-0.0081	0.1006	-0.1272	0.1442
com2	-0.2779	0.0777	0.0524	-0.3456	0.0981	-0.0052	-0.0683	-0.0042	-0.0717	-0.0356	-0.0168
com3	-0.0884	0.0093	-0.0619	0.5488	0.0487	0.0430	-0.1765	-0.0083	0.0413	-0.0901	0.1631
ABC	0.0513	-0.1968	-0.1882	0.3596	-0.1619	0.0685	0.3778	0.0222	-0.0407	-0.1050	0.0140
directed_com	0.3743	0.0638	0.1126	-0.2705	-0.0336	-0.0403	-0.0129	0.0035	-0.0173	0.2439	-0.2226

	BC	Prairie	ON	Quebec	Atlantic	com1	com2	com3	ABC	direct~m
BC	1.0000									
Prairie	-0.2335	1.0000								
ON	-0.4251	-0.2854	1.0000							
Quebec	-0.3126	-0.2098	-0.3821	1.0000						
Atlantic	-0.1283	-0.0861	-0.1568	-0.1153	1.0000					
com1	0.0289	-0.0183	-0.1224	0.0634	0.1225	1.0000				
com2	-0.1187	-0.0622	0.0692	0.1148	-0.0342	-0.1790	1.0000			
com3	-0.0186	0.1165	-0.0759	-0.0091	0.0387	-0.1825	-0.2653	1.0000		
ABC	0.1050	0.0969	-0.1070	-0.0369	-0.0630	-0.1017	-0.1478	-0.1507	1.0000	
directed_com	0.0380	-0.0874	0.1432	-0.1079	-0.0480	-0.2827	-0.4109	-0.4190	-0.2334	1.0000

Figure 5.1 Correlations matrix for independent variables

Multicolinearity is a risk in these kinds of analysis. If one or more of the independent variables are significantly correlated with each other, it would not necessarily reduce the overall explanatory power (R^2) of a regression but it might significantly change the assigned impact of the explanatory power of the independent coefficients.

For the 156 examples, the t –stat which matches 90%, 95%, 99% significance level is as the following table 5.1.

If the correlation coefficient exceeds certain number shown in the table, then there is the potential for multicolinearity.

Table 5.1 Critical values for significant correlations

dF=155	Significant	T	Correlation coefficient
	90%	1.65	0.132
	95%	1.98	0.157
	99%	2.61	0.206

Source: Author's calculations

The regional dummies for BC and Ontario have a correlation coefficient of -0.4251, which means that the two variables are significant negatively correlated. We have controlled for this by leaving the Ontario dummy out of the regression.

The fishery dummy is also significant positively correlated with the Atlantic region (+0.4205). The reason is that the activity related to the fisheries is too small (with only 3 projects) and almost half of the fishery program is in Atlantic.

The solution chosen was to combine fisheries with environment. Removing those two variables from the analysis solves most of the significant correlations (see figure 5.2).

```
. correlate   PI research international INTERGE3LS health agriculture environment forestry tech GE3LS BC P
> rairie ON Quebec Atlantic com1 com2 com3 ABC directed_com
(obs=155)
```

	PI	research	intern~l	INTERG~S	health	agricu~e	enviro~t	forestry	tech	GE3LS
PI	1.0000									
research	0.1820	1.0000								
internatio~l	0.1935	0.1319	1.0000							
INTERGE3LS	0.0251	-0.1569	-0.1764	1.0000						
health	0.1338	0.1825	0.1320	-0.0125	1.0000					
agriculture	-0.0944	-0.1353	-0.1171	0.1009	-0.3469	1.0000				
environment	0.0173	-0.2891	-0.0706	0.0787	-0.3961	-0.1223	1.0000			
forestry	-0.0780	-0.1039	-0.0556	0.2120	-0.2783	-0.0860	-0.0982	1.0000		
tech	0.0262	0.1416	0.0290	-0.1640	-0.3842	-0.1186	-0.1355	-0.0952	1.0000	
GE3LS	-0.0866	0.0930	-0.0145	-0.1907	-0.2426	-0.0905	-0.1033	-0.0726	-0.1002	1.0000
BC	-0.1086	-0.1769	0.0687	0.1292	-0.0343	-0.0434	0.0044	0.2052	-0.0757	0.0093
Prairie	-0.1013	-0.1879	-0.0921	0.0898	-0.2685	0.3805	0.0245	-0.1040	0.0330	0.1108
ON	0.2368	0.3905	0.0974	-0.1483	0.1079	-0.1439	-0.0621	-0.1339	0.1632	-0.0403
Quebec	-0.0591	-0.0072	-0.0713	-0.0656	0.1566	-0.1208	-0.0555	0.0512	-0.0948	-0.0251
Atlantic	-0.0274	-0.1952	-0.0733	0.0493	-0.0438	0.0339	0.2029	-0.0571	-0.0788	-0.0601
com1	-0.1534	-0.0442	-0.0020	-0.2422	0.0003	-0.0451	0.0577	-0.0081	-0.1272	0.1442
com2	-0.2779	0.0777	0.0524	-0.3520	0.0981	-0.0052	-0.0934	-0.0042	-0.0356	-0.0168
com3	-0.0884	0.0093	-0.0619	0.4840	0.0487	0.0430	-0.1463	-0.0083	-0.0901	0.1631
ABC	0.0513	-0.1968	-0.1882	0.3681	-0.1619	0.0685	0.3334	0.0222	-0.1050	0.0140
directed_com	0.3743	0.0638	0.1126	-0.1604	-0.0336	-0.0403	-0.0192	0.0035	0.2439	-0.2226

	BC	Prairie	ON	Quebec	Atlantic	com1	com2	com3	ABC	direct~m
BC	1.0000									
Prairie	-0.2335	1.0000								
ON	-0.4251	-0.2854	1.0000							
Quebec	-0.3126	-0.2098	-0.3821	1.0000						
Atlantic	-0.1283	-0.0861	-0.1568	-0.1153	1.0000					
com1	0.0289	-0.0183	-0.1224	0.0634	0.1225	1.0000				
com2	-0.1187	-0.0622	0.0692	0.1148	-0.0342	-0.1790	1.0000			
com3	-0.0186	0.1165	-0.0759	-0.0091	0.0387	-0.1825	-0.2653	1.0000		
ABC	0.1050	0.0969	-0.1070	-0.0369	-0.0630	-0.1017	-0.1478	-0.1507	1.0000	
directed_com	0.0380	-0.0874	0.1432	-0.1079	-0.0480	-0.2827	-0.4109	-0.4190	-0.2334	1.0000

Figure 5.2 Correlation matrix for independent variables (fixed data)

The further matrix is made under the estimated model which only show the used dummy and variables, chosen was under the logic of the final regression (Model D in table 5.4).

```
. correlate   PI research international INTERGE3LS health ON directed_com
(obs=155)
```

	PI	research	intern~l	INTERG~S	health	ON	direct~m
PI	1.0000						
research	0.1820	1.0000					
internatio~l	0.1935	0.1319	1.0000				
INTERGE3LS	0.0251	-0.1569	-0.1764	1.0000			
health	0.1338	0.1825	0.1320	-0.0125	1.0000		
ON	0.2368	0.3905	0.0974	-0.1483	0.1079	1.0000	
directed_com	0.3743	0.0638	0.1126	-0.1604	-0.0336	0.1432	1.0000

Figure 5.3 Correlation matrix for independent variables (model D in table 5.4)

5.2. **The Basic OLS & Model Building**

At this point it would be a good idea to see the structure of the models that are evaluated (see table 5.3), and the summary statistics which has been used in the following OLS.

Table 5.3 Independent Variable Description

Independent variable		
Number	Category	Variable
VAR1	Leadership	PI
VAR2	Investment	Maclean research index
VAR3	Partnership	International co-funding
VAR4	GC strategy	interGE3LS
VAR5	Sector	health
VAR6		agriculture
VAR7		environment
VAR8		forestry
VAR9		Technology
VAR10		GE3LS
VAR11	Competition	com1
VAR12		com2
VAR13		com3
VAR14		ABC
VAR15		Directed
VAR16	Region	BC
VAR17		Prairie
VAR18		ON
VAR19		Quebec
VAR20		Atlantic
R^2 (%) for regression Y_1 GC-total, Y_2 open-com		
N=156		

Variable	Obs	Mean	Std. Dev.	Min	Max
GC_total	156	.6413462	.510989	.02	2.64
open_com	95	1.052211	.6634102	.12	2.84
PI	156	10.04981	9.573183	.2	53
research	156	2.511923	.739311	.43	3.51
internatio~l	155	.2967742	.458317	0	1
INTERGE3LS	156	.3205128	.4681767	0	1
health	156	.525641	.5009503	0	1
ON	156	.3397436	.4751474	0	1
directed_com	156	.3910256	.4895517	0	1

Figure 5.4 Summary Statistics

In effect, we test a number of configurations of consolidating or unpacking various dummy options to find the best fit. All of the regressions include the core independent

variables; the PI-HI measure, the Maclean's ranking, the dummy for GE3LS and the technology variable.

The models are designed under the logic after table 5.3. Apart from the four core independent variables, the PI-HI measure, the Maclean's research ranking index, international co-funding, interGE3LS for the strategy, the dummy is added in an order to see the changed R^2 of the process.

Table 5.4 Multivariate Model building

Independent variables	Model A	Model B	Model C	Model D
Leadership indicators: PI	√	√	√	√
Investment indicators: Maclean Research index	√	√	√	√
Partnership indicators: International-co-funding	√	√	√	√
GC Strategy indicators: INTER-GE3LS	√	√	√	√
Dummies for SECTOR: Health		√	√	√
Dummies for COMPETITION: Directed			√	√
Regional dummies ON				√

5.3. **Regression of Y-GC total**

We can see that as we expand the scope of dummies, the overall model fit increases and a larger share of the allocation of funds is explained by the evidence available at the time of the decisions. In this sense, the model helps to quantify the relationship between the goals and allocations of Genome Canada.

Table 5.5 presents the results of estimating OLS with Y-GC total as the dependent variable.

Four separate regressions are presented; others with more dummy variables were estimated but they did not improve the fit and are not included here.

Table 5.5 OLS estimation result on Y-GC total (Detailed table see Appendix V)

Dependent Variable Y-GC total				
Independent Variable	Model A	Model B	Model C	Model D
Intercept	0.21	0.16	0.18	0.32**
Leadership indicators: PI	0.01***	0.01***	0.01***	0.02***
Investment indicators: Maclean research index	0.06	0.03	0.02	0.007
Partnership indicators: International co-funding	0.23***	0.2**	0.2**	0.21***
GC Strategy indicators: Inter-GE3LS	0.26***	0.25***	0.26***	0.19**
Dummies-SECTOR: Health		0.3***	0.3***	0.26***
Dummies-REGION: ON			0.05	0.07
Dummies -COMPETITION: Directed				-0.41***
Number of observation	155	155	155	155
F Statistics	7.75	9.97	8.33	12.80
Adjusted R^2	0.15	0.23	0.22	0.35
Significance levels (*p* value): * p<0.1; ** p<0.05; *** p<0.01				

Model D fit the highest R^2. The basic equation which follows the objectives stated in GC in Regression Y-GC total is as following:

Y=0.32 +0.02*(PI) +0.007*(Research) +0.21*(International) +0.19*(INTERGE3LS)
2.40 4.69*** 0.14 2.77*** 2.51****
+0.26*(Sector-health) +0.07*(Region-ON) -0.41*(Directed)
3.79* 0.92 5.46*** (5.3.1)**

We will interpret the result based on Model D, but also discuss results of the other models.

We see that the intercept term is equal to 0.32, which means the funding share of a project in total fund pool of competitions when the value of all other independent variables are equal to zero would be 0.32% (significant at 95% level).

Moreover, on average, a project's Principal Investigator (PI) reputation, measured by the HI index, increases the project share by 0.02% for each unit increase index in HI (significant at 99.9% level), other things being equal.

The host institution also has little effect. On average, the share of GC contribution to each project in the total pool of all Genome Canada contributions will increase 0.007 for each additional index point (not significant). The project's host institution of research capability index is measured by total research dollars per full time faculty member (10000$). On average, projects with international co-founding share approximately 0.21% (99% confidence level) more than a project which has matching funds only from domestic sources, other things being equal.

Moreover, on average, an INTER-GE3LS project is expected to have approximately 0.19% (95% confidence level) more than a project which is not, other things being equal.

Moving on to the coefficient for sector, on average a health project is expected to have approximately 0.26% (99.9% confidence level) more than a project which is not, other things being equal.

For the Region dummy, on average, a project in Ontario is expected to have approximately 0.07% (not significant) more than projects which are not in Ontario, other things being equal. In short, there is no special regional bias.

A project which is not from Com I, II, II and ABC (i.e. directed-com) is expected to share approximately 0.41% (99.9% confidence level) less than an open-competition project, all other things being equal. In short, the open competition grants were larger.

Model D contains more detailed dummy variables, such as the regional dummies, the sector dummies and the competition dummies as the adj-R^2 reaches up to 35% for these regressions.

Other more specified models were calculated but the adj-R2 did not improve measurable. Given Occam's razor that the simplest explanation that explains the most is best, and the principles of parsimony, economy and succinctness, Model D was chosen, as it used the least variables to explain the most.

5.4. **Regression of Y-open com**

Three of the many regressions attempted are reported here. Those with more dummies were rejected as they did not materially improve the fit.

Since the Y-open com regression is only about the open review process, the competition dummy is not suitable to test in this section.

Table 5.6 OLS estimation result on Y-open com

Dependent Variable Y-open com			
Independent Variable	Model A	Model B	Model C
Intercept	0.57**	0.52**	0.65***
Leadership indicators: PI	0.02*	0.02	0.02*
Investment indicators: Maclean rank	0.07	-0.00	-0.09
Partnership indicators: International co-funding	0.27*	0.21	0.20
GC Strategy indicators: Inter-GE3LS	0.26*	0.29**	0.28**
Dummies-SECTOR: Health		0.47***	0.48***
Dummies-REGION: ON			0.30*
Number of observation	94	94	94
F Statistics	3.27	5.65	5.47
Adjusted R^2	0.15	0.20	0.22
Significance levels (*p* value): * p<0.1; ** p<0.05; *** p<0.01			

The model D fit the highest R^2. The basic equation which follows the objectives stated in GC in Regression Y-open com is as following:

Y= 0.65 + 0.02 * (PI) - 0.09 * (Research) + 0.2 * (International)

2.74* 1.77* 0.97 1.44**

+0.28 * (INTERGE3LS) +0.48 * (Sector-health) + 0.3 * (Region-ON)

2.10 3.81*** 1.92***

(5.4.2)

We see that the intercept term is equal to 0.65, which means the funding share of a project in open pool of competitions (I, II, III, ABC) when the value of all other independent variables are equal to zero would be 0.65% (significant at 99% level).

Leadership continues to matter. On average, the quality of a project's Principal Investigator (PI), which is measured by HI index of the lead-person, would share 0.02% more of the funding share of a project in open pool of competitions (I, II, III, ABC) (90% confidence level) for each unit increase index in HI, other things being equal.

International co-funding, on average, improves a project's budget share by 0.2% (not statistically significant) more than a project which is only supported from domestic source, other things being equal.

The host institution also has little effect. On average, the GC contribution to each project in the open competitions increases 0.09 for each additional index point (not significant). The project's host institution of research capability index is measured by research funding per full-time faculty member (10000$).

However, on average, an INTERGE3LS project is expected to have approximately 0.28% (95% confidence level) more than a project which is not, other things being equal.

Moreover, for the coefficient for sector, on average, a health project is expected to have approximately 0.48% (99.9% confidence level) more than a project which is not, other things being equal.

On average, each ON project is expected to have approximately 0.3% (90% confidence level) more share of GC contribution in the open fund pool of all Genome Canada contribution than a project which is not, other things being equal. This suggests that the peer reviewers appear to be more influenced by the location of the project than Genome Canada staff.

Model C, contains the regional dummies and the sector dummies. The adj-R^2 reaches a peak at 22%; more specified models with other contextual variables were tested but they offer little additional explanation power (based on the static adjusted R^2).

Overall, this model suggests the processes in Competitions I, II, III and ABC delivered a weaker fit with the strategic of objectives of Genome Canada than the processes used by Genome Canada staff to develop the directed projects. This may be an artifact of the lessons learned from the earlier open competitions that were applied to the directed investments. However, there is some possibility that there may have been cognitive biases operating in the open competitions, as the dummy for the Ontario region is positive and significant at 90% level, which should not be observed in a competition where research excellence is the goal rather than allocations based on past capacity.

6. Summary & Policy Implications

6.1. **Summary**

This study has added to the policy evaluation literature, offering specific insights into evaluation of Genome Canada. GC was established in April, 2000 to provide funding and information resources related to genomics research. GC research targets many key areas, such as health, agriculture, environment, forestry, fisheries, energy and mining.

Since then, the scientific community has partnered with government, the private sector, and international organizations to fund research projects on genomics related subjects. Four open competitions (I, II, III and Applied Genomics in Bio-products and Crops or ABC), combined with a wide array of internally targeted and developed projects, have collectively been allocated more than C$2 billion in total investment for the 2000-2014 period.

This study assesses how well these research projects fit the stated goals of Genome Canada. The study assesses the fit between the goals and research investment decisions of GC. As a first step in this research, we conducted a review of Genome Canada operations to develop the background understanding of the system and its structure. After reviewing the goals, structure, selection processes and progress reports, we found that there was no explicit assessment of the fit between the stated goals and resource allocation decisions. This study targeted to fill this gap.

Second, we investigated the methods used by GC to develop and implement their goals. Once we understood these methods, we developed a research approach to examine the fit between the goals and the outputs. We explored the resource allocation decisions of GC, especially, the individual projects from different sectors. An econometric model was built to test the allocations of funding for projects against the overall program stated objectives, namely to: develop and implement a coordinated strategy for the technology in Canada; bring together industry, governments, universities, research hospitals and the public to support large-scale genomics and proteomics research projects; provide accessibility to

science and technology platforms to researchers; and assist in attracting co-funding for projects from both domestic and international investors.

Third, we determined that the review processes contain scientific, financial and management criteria. By using the STATA tool, we tested the relationship between the share of funds allocated to specific projects in the competitions and in the directed investments and the stated goals of the organization. The analysis revealed that the overall fit for the entire investment program between 2001 and 2011 was about 34%, which is quite strong. We found the most important variable affecting resource allocation was the quality of the principal investigator. Other stated goals of GC were either less important or insignificant. By segmenting the analysis into the open competition investments alone, we discovered the fit deteriorated (R^2 dropped from 34% to 22%), which suggests the directed investments are a stronger fit with the goals. While we could not conclusively determine the cause, it might be attributed to (1) weaknesses in the peer-review processes involving a large number of competitive projects, (2) greater competence in adjudication as the directed investments mostly followed the four open competitions, or (3) it could be due to particularly effective and strategic effort by Genome Canada staff. Further analysis would be needed to determine this.

6.2. **Conclusions**

First, the results of our study shows that about up to 35% of the variance in funding by project can be explained by goals of GC. This is actually quite good for this type of program.

Second, the key variables that seemed to influence allocations were: health, ON, PI, competitions I, II, III, ABC, research, GE3LS, INTERGE3LS and International co-founders.

Third, somewhat surprisingly the fit for the open competitions was not as strong as for the entire portfolio. By inference, this means that the allocations directed by Genome Canada staff (i.e. not engaged in open competition) were generally more strategic (keep

in mind we cannot confirm in this study that their outputs and outcomes were any different—that would be a different type of analysis).

This may be surprising to many, as there is a general view that bureaucrats are more susceptible to political interference than arms-length openly competitive processes. One of two factors could be contributing to this divergence. It is possible that the competitive process triggers cognitive gaps and biases among the peer-reviewers. There is some theory and evidence that peer review systems that are directed to assess multiple projects over a diverse set of variables will revert to system 1, fast and intuitive thinking that would lead to anchoring on a few operative factors and satisficing activity (Kahneman 2002)[31]. Whether that is working here could be examined experimentally. The differential importance of sector and region for peer reviewers suggests something is going on here. Alternatively, it may be that the staffs of Genome Canada and the regional genome centers are as susceptible to incentives as many might hypothesize, but that their incentives drive them to proactively backfill and compensate for any gaps in the open competition results. It would be necessary to look at the incentive and operational mandates of the Genome Canada staff to determine what drives these behaviors.

6.3. **Limitations**

This study was done using publicly available data. Access to internal Genome Canada data—including the detailed proposals for the projects—would allow us to calibrate the model more precisely and, in a perfect world, determine if there are any learning by doing effects as the organization has matured.

A second limitation is that we do not have any counterfactuals. The share of allocations was used as an in-sample differentiator. In a perfect world we would have full access to the structure and details of those proposals that failed to advance from LOI to full proposal and that were not funded. That would provide an all-in analysis of the efficacy and fit of the Genome decision system relative to its stated goals.

6.4. **Extensions**

This study raises two interesting possibilities for further work. First, pending access to more detailed data on both successful and unsuccessful projects, it should be possible to more effectively refine the model and isolate the effect of key variables in decision making.

This then could be used to assess the effect of framing and choice architecture in research decision making. As noted above, this analysis tends to provide empirical evidence in support of the possibility that peer-evaluation systems are cognitively limited in the context open competitions. We believe experimental work specifically related to the choices facing the peer-reviewers in Genome Canada could help more effectively develop appropriate choice architecture.

APPENDIXES

Appendix I: PI kurtosis & codebook

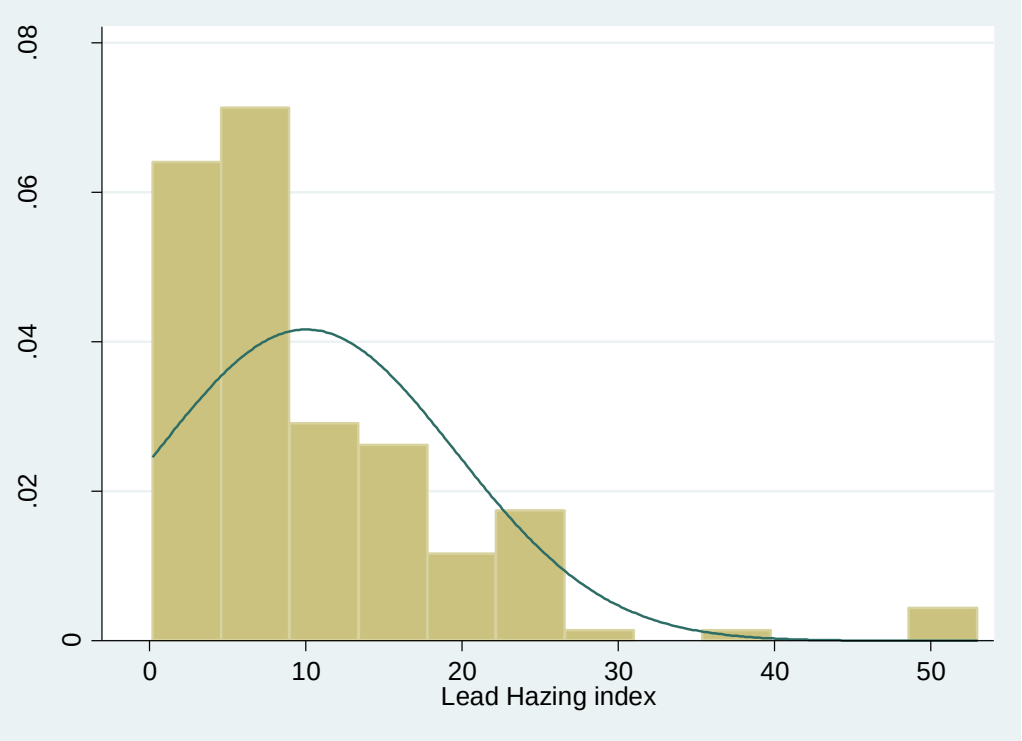


Figure 1 PI kurtosis

PI

type:	numeric (**float**)				
range:	[.2,53]		units:	.01	
unique values:	130		missing .:	0/156	
mean:	10.0498				
std. dev:	9.57318				
percentiles:	10%	25%	50%	75%	90%
	1.19	3.775	7.35	13.925	23.16

Figure 2 PI codebook

Appendix II: Maclean Ranking

Table 3 Maclean Ranking

University	Total Research Dollars ($ per full-time faculty member)	Total Research Dollars (10000$ per full-time faculty member)
1. Toronto	350995	3.51
2. Alberta	309332	3.09
3. McMaster	308605	3.09
4. McGill	268730	2.69
5. Montréal	257238	2.57
6. UBC	238875	2.39
7. Queen's	216764	2.17
8. Laval	211253	2.11
9. Ottawa	194084	1.94
10. Guelph	191884	1.92
11. Manitoba	175400	1.75
12. Western	171784	1.72
13. Calgary	169787	1.70
14. Waterloo	162683	1.63
15. Victoria	158087	1.58
16. Saskatchewan	156464	1.56
17. Dalhousie	131691	1.32
18. Carleton	101464	1.01
19. Simon Fraser	99452	0.99
20. UNBC	98700	0.99
21. Sherbrooke	97811	0.98
22. New Brunswick	91701	0.92
23. Memorial	81761	0.82
24. UPEI	71419	0.71
25. Windsor	66923	0.67
26. UQAM	65824	0.66
27. Lakehead	64683	0.65
28. UOIT	63601	0.64
29. Trent	52902	0.53
30. Regina	52893	0.53
31. York	48195	0.48
32. Lethbridge	47068	0.47
33. Laurentian	46541	0.47
34. St. Francis Xavier	45688	0.46
35. Concordia	43483	0.43
36. Cape Breton	40077	0.40
37. Saint Mary's	35446	0.35

38. Moncton	30752	0.31
39. Ryerson	30587	0.31
40. Winnipeg	25743	0.26
41. Acadia	25530	0.26
43. Mount Allison	23956	0.24
42. Mount Saint Vincent	24028	0.24
44. Brock	23636	0.24
45. Wilfrid Laurier	19620	0.20
46. Brandon	14528	0.15
47. Nipissing	14090	0.14
48. Bishop's	9054	0.09
49. St. Thomas	6941	0.07

Source: http://tools.macleans.ca/ranking2008/selectindicators.aspx

Appendix III: Calculation of Frequency

research — Maclean Total Researc

type:	numeric (float)				
range:	[.43,3.51]		units:	.01	
unique values:	25		missing .:	0/156	
mean:	2.51192				
std. dev:	.739311				
percentiles:	10%	25%	50%	75%	90%
	1.56	1.94	2.39	3.09	3.51

health

type:	numeric (float)			
range:	[0,1]		units:	1
unique values:	2		missing .:	0/156
tabulation:	Freq.	Value		
	74	0		
	82	1		

agriculture

type:	numeric (float)			
range:	[0,1]		units:	1
unique values:	2		missing .:	0/156
tabulation:	Freq.	Value		
	141	0		
	15	1		

environment

type:	numeric (float)			
range:	[0,1]		units:	1
unique values:	2		missing .:	0/156
tabulation:	Freq.	Value		
	140	0		
	16	1		

forestry

type:	numeric (**float**)		
range:	[**0**,**1**]	units:	**1**
unique values:	**2**	missing .:	**0**/**156**

tabulation:

Freq.	Value
145	**0**
11	**1**

fisheries

type:	numeric (**float**)		
range:	[**0**,**1**]	units:	**1**
unique values:	**2**	missing .:	**0**/**156**

tabulation:

Freq.	Value
153	**0**
3	**1**

tech

type:	numeric (**float**)		
range:	[**0**,**1**]	units:	**1**
unique values:	**2**	missing .:	**0**/**156**

tabulation:

Freq.	Value
138	**0**
18	**1**

GE3LS

type:	numeric (**float**)		
range:	[**0**,**1**]	units:	**1**
unique values:	**2**	missing .:	**0**/**156**

Appendix IV: Table of Key Variables

N=156

PI	research	international	INTERGE3LS	health	ON	Directed
0.67	3.51	0	0	0	1	0
1.45	1.56	1	0	0	0	0
16.11	2.51	0	1	1	1	1
5.68	1.75	0	1	0	1	0
3.79	1.72	0	1	0	0	0
7.95	1.56	0	1	0	0	0
13.73	2.69	0	1	0	0	0
16.11	1.56	0	1	1	1	1
0.64	0.92	0	0	0	0	0
5.08	3.09	0	1	0	0	0
5.15	2.39	0	1	0	0	0
37.72	3.09	1	0	1	1	1
5.15	2.39	1	0	0	0	1
2.44	3.09	0	0	0	0	0
3.76	2.33	0	1	1	0	1
2.57	1.56	0	1	0	0	0
12.25	3.51	0	1	0	1	0
0.5	3.51	0	0	0	0	1
10.94	1.94	0	0	0	0	1
9.08	3.51	0	0	0	1	1
1.26	1.75	0	0	0	1	1
2.22	3.51	0	0	0	1	0
5.68	3.51	1	0	0	0	1
12	3.51	1	0	0	1	1
5.49	2.17	1	0	0	0	0
13.59	2.39	0	0	0	1	1
0.25	3.51	1	0	0	1	0
7.41	1.58	0	0	0	0	1
25.13	1.92	0	0	0	1	1
7.08	2.69	1	0	0	0	1
6.62	1.7	0	1	0	0	0
15.91	3.51	1	0	0	1	1
5.38	2.57	0	0	0	0	1
17.52	2.39	0	0	0	1	1
24.08	2.39	0	1	0	0	0
1.88	2.39	0	0	0	0	0
11.54	2.57	0	1	0	0	1
13.26	1.7	0	1	0	1	0

20.26	1.7	0	1	0	0	0
25.13	1.92	0	0	0	1	1
4.76	1.92	0	1	0	0	1
25.13	1.92	1	0	0	1	1
3	1.32	0	0	0	0	0
15.57	1.75	0	1	0	0	0
10.99	2.39	0	1	0	0	1
25.13	1.92	0	1	0	0	0
12.45	2.39	0	1	0	0	1
1.39	0.43	1	0	0	0	0
1.19	3.09	0	0	0	1	0
26.11	2.39	0	0	0	1	1
6.13	0.43	0	1	1	0	0
3.91	0.99	0	0	0	0	0
6	0.92	1	0	0	0	1
2.64	2.39	0	1	0	0	0
22	1.58	1	1	1	0	0
1.94	2.11	1	1	0	0	0
7.68	0.92		0	0	0	0
6.15	2.74	0	1	0	0	1
1.8	2.39	1	0	0	0	0
7.89	2.74	0	1	1	0	0
19.69	2.25	0	1	0	0	1
10.18	2.39	0	0	0	0	1
8	2.39	0	0	0	0	1
2	2.57	0	0	0	0	0
1.38	1.56	0	0	0	1	0
11.39	2.39	0	1	0	0	0
6.63	3.51	1	0	0	1	0
6.94	1.69	0	0	0	0	0
8.4	2.57	0	0	0	0	0
3.43	1.7	0	0	0	0	0
7.68	2.39	0	0	0	0	0
1.33	3.09	0	0	0	0	0
17.65	3.51	0	0	1	0	0
7.76	2.57	0	0	0	0	0
6.13	3.51	1	0	0	1	0
5.14	2.39	0	0	0	0	0
6.63	3.51	1	0	0	1	0
8	3.51	1	0	1	1	1
7.53	2.57	0	0	1	0	0
10.95	3.51	1	0	1	1	0
1.09	1.56	1	0	1	0	0

15.5	2.39	0	0	1	1	1
19.57	2.69	0	0	1	0	0
5.44	2.69	0	0	1	0	0
1.78	2.39	1	0	1	0	0
0.4	2.57	0	0	1	0	0
15.22	2.39	1	0	1	0	1
4.52	2.69	1	0	1	0	0
12.69	3.51	0	0	1	1	0
11.52	2.69	1	0	1	0	0
1.58	2.39	0	1	1	0	0
6.25	3.51	0	1	1	1	0
53	3.51	1	0	1	1	1
24	2.69	0	0	1	0	1
9.1	2.39	0	0	1	0	0
3.23	2.69	0	0	1	0	0
24	2.69	0	0	1	0	1
11.09	3.51	0	1	1	0	1
24	2.69	1	0	1	0	1
0.7	3.51	0	0	1	1	0
0.25	1.94	0	0	1	0	1
8.21	2.39	0	0	1	0	1
5.36	2.69	1	0	1	0	1
0.7	2.39	0	0	1	0	1
19.24	3.51	0	0	0	0	1
6.34	2.39	0	1	1	0	0
0.5	0.99	1	0	1	0	0
5.44	0.98	0	0	1	0	0
4.17	1.72	0	0	0	0	1
53	3.51	1	0	1	1	1
0.6	1.75	0	1	1	1	0
8.89	2.39	1	0	1	0	0
5.19	2.69	0	0	1	0	0
0.28	2.57	0	0	1	0	0
1.32	2.11	0	0	1	0	1
18.72	3.51	1	0	1	1	1
2.58	2.39	1	1	1	0	0
22.98	2.39	1	0	1	0	1
5.36	2.69	0	0	1	0	0
15.09	1.94	0	0	1	1	0
13.57	3.51	0	0	0	0	1
5.61	3.51	1	1	1	1	1
0.2	3.51	0	0	1	1	0
24.5	1.72	1	1	1	1	0

14.12	3.51	0	0	1	1	0
14.19	3.09	1	0	1	1	1
12.3	2.69	0	0	1	0	0
7.35	3.51	0	0	1	1	0
3.85	3.51	0	1	1	1	0
4.9	2.39	0	0	1	0	0
10.27	1.32	0	1	1	0	1
2.5	2.57	0	0	1	0	0
53	3.51	1	0	1	1	1
10	1.94	0	0	1	1	1
14.13	2.57	0	1	1	0	0
5.08	3.51	0	0	1	1	0
7.35	3.51	0	1	1	1	0
5.13	2.57	0	0	1	0	0
6.78	3.51	0	0	1	1	0
13.59	2.39	0	1	0	0	0
7.56	2.39	0	0	1	1	1
1.45	2.39	1	1	1	0	0
17.51	2.69	1	1	1	0	0
5.08	3.3	0	1	1	1	0
15.45	3.09	1	0	0	0	1
23.16	2.57	0	1	1	0	0
27.92	3.51	1	0	1	1	1
12.6	2.39	1	1	1	0	1
18.31	3.51	0	1	1	0	0
10	1.94	0	0	1	1	1
1.09	2.39	1	1	1	0	0
8.97	2.69	0	0	1	0	0
0.82	2.39	0	0	1	1	1
19.04	2.39	0	1	1	0	1
5.09	3.51	1	0	1	1	0
5.24	2.39	1	0	1	0	0

Appendix V: Original STATA DATA

	projecttitle	GC_total	open_com	PI	research
1	Functional genomics of Arabidopsis	.12	.17	.67	3.51
2	Functional genomics of abiotic stress in crops	1.43	2.01	1.45	1.56
3	Whole Genome Selection through Wide Imputation in Beef Cattle	.57	.	16.11	2.51
4	Value Addition to Genomics and GE3LS (VALGEN)	.37	.53	5.68	1.75
5	Genomics in Agricultural Pest Management (GAP-M)	.41	.58	3.79	1.72
6	Total Utilization Flax GENomics (TUFGEN)	.83	1.16	7.95	1.56
7	Bridging comparative, population and functional genomics to identify and experimentally validate novel regulatory regions and genes for crop improvement	.32	.45	13.73	2.69
8	CTAG-Canadian Triticum Advancement through Genomics	.6	.	16.11	1.56
9	Canadian potato genome project	.28	.39	.64	.92
10	Designing oilseeds for tomorrow's market	.98	1.38	5.08	3.09
11	Grape and Wine Genomics	.24	.34	5.15	2.39
12	Bovine genome sequencing program: Full-length cDNA sequencing	.89	.	37.72	3.09
13	GrapeGen - A genomic approach to the identification of the genetic and environmental components underlying berry quality in grapevine	.46	.	5.15	2.39
14	Enhancing canola through genomics	.53	.75	2.44	3.09
15	Application of genomics to improve swine health and welfare	.72	.	3.76	2.33
16	Crop Adaptation Genomics - Use of genomic tools for crop improvements in temperate climates	.59	.83	2.57	1.56
17	Quantum dot diagnostics: Simultaneous genomic and proteomic profiling of multiple pathogens at point-of-care	.71	1	12.25	3.51
18	Technologies for Methylome Studies	.06	.	.5	3.51
19	Proteomic Technologies for the Study of Rare Cells	.11	.	10.94	1.94
20	Multiplexed MicroRNA Detection on an Electronic Chip	.07	.	9.08	3.51
21	Enabling Technologies for Embryonic Stem Cell Functional Genomics	.05	.	1.26	1.75
22	Development and applications of functional genomics technologies	.65	.91	2.22	3.51
23	Software Tools to Simplify Gene Function Prediction	.14	.	5.68	3.51
24	Massively Multiparametric Flow Cytometer Analyzer	.21	.	12	3.51
25	Development of enabling technologies for proteomic research	1.25	1.77	5.49	2.17
26	Production-scale Deployment of Next-generation Sequencing Instruments	.14	.	13.59	2.39
27	Fiber optic nucleic acid biosensor based gene profiling: Proof of principle by screening for drug leads for orphan neurodegenerative disorders and SNP analysis	.22	.31	.25	3.51
28	MS-based Structural Proteomics for Drug Development and Design	.07	.	7.41	1.58
29	Environmental Barcoding through Massively Parallelized Sequencing	.12	.	25.13	1.92
30	Integrated Proteomics Platforms for High-throughput Biomarker Discovery and Validation	.11	.	7.08	2.69
31	Four-dimensional modelling of genetic disease patterns	.33	.47	6.62	1.7
32	Automated Three-dimensional Phenotyping of Mouse Embryos	.1	.	15.91	3.51
33	High-throughput, High-dimentional, Multi-parametric Analysis of the Immune System	.05	.	5.38	2.57

	internatio~l	INTERGE3LS	health	agriculture	environment	forestry	tech	GE3LS	BC	Prairie	ON	Quebec	Atlantic	com1	com2	com3	ABC	directed_com
1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0
2	1	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0
3	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
4	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0
5	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0
6	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0
7	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0
8	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
9	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0
10	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
11	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0
12	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
13	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
14	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0
15	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
16	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
17	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0
18	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
19	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
20	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
21	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
22	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
23	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
24	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
25	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0
26	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
27	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
28	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1
29	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
30	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1
31	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
32	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
33	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1

	projecttitle	GC_total	open_com	PI	research
34	Towards Single Cell Genomics	.13	.	17.52	2.39
35	Genomics of Sunflower	.73	1.02	24.08	2.39
36	Microbial envirogenomics: micro-organisms and their interaction with the environment	.34	.48	1.88	2.39
37	Improving Bioremediation of polluted soils through Environmental Genomics	.56	.	11.54	2.57
38	Metagenomics for Greener Production and Extraction of Hydrocarbon Energy	.74	1.04	13.26	1.7
39	Synthetic Biosystems for the Production of High-Value Plant Metabolites	.94	1.33	20.26	1.7
40	International Barcode of Life Project (iBOL)	.02	.	25.13	1.92
41	Biomonitoring 2.0: A high-throughput genomics approach for comprehensive biological assessment of environmental change	.23	.	4.76	1.92
42	International Barcode of Life Project (iBOL)	1.26	.	25.13	1.92
43	Understanding prokaryotic genome evolution and diversity	.29	.41	3	1.32
44	Microbial Genomics for Biofuels and Co-products from Biorefining Processes	.71	1.01	15.57	1.75
45	Next-generation integrated pest management tools for beekeeping	.42	.	10.99	2.39
46	Canadian barcode of life network	.8	1.13	25.13	1.92
47	Applied Metagenomics of the Watershed Microbiome	.23	.	12.45	2.39
48	Genomic approaches to identify fungal enzymes for industrial and environmental applications	.55	.78	1.39	.43
49	Genomic analyses of soil micro-organisms	.43	.6	1.19	3.09
50	POPCAN: Genetic improvement of poplar trees as a Canadian feedstock	.71	.	26.11	2.39
51	Genozymes for Bioproducts and Bioprocesses Development	1.19	1.68	6.13	.43
52	Genomics research on atlantic salmon project (GRASP)	.45	.64	3.91	.99
53	Pleurogene - Flatfish genomics: Enhancing commercial culture of Atlantic halibut and Senegal sole	.29	.	6	.92
54	Atlantic cod genomics and broodstock development	1.28	1.81	2.64	2.39
55	Consortium for genomic research on all salmonids project (cGRASP)	1.09	1.53	22	1.58
56	Arborea II: Genomics for molecular breeding in softwood trees	1.02	1.44	1.94	2.11
57	Comparative structural and functional spruce genomics	.12	.17	7.68	.92
58	AdapTree: Assessing the adaptive portfolio of reforestation stocks for future climates	.34	.	6.15	2.74
59	Forestry genomics: Mechanisms of wood formation and pest resistance in forest trees using spruce, poplar and arabidopsis	.78	1.1	1.8	2.39
60	Genomics-Enhanced Forecasting Tools to Secure Canada's Near-Term Lignocellulosic Feedstock Supply for Bioenergy using the Mountain Pine Beetle-Pinus spp. System	.54	.76	7.89	2.74
61	SMarTForest : Spruce Marker Technologies for Sustainable Forestry	.72	.	19.69	2.25
62	Genomics-Based Forest Health Diagnostics and Monitoring	.3	.	10.18	2.39
63	Harnessing microbial diversity for sustainable use of forest biomass resources	.57	.	8	2.39
64	Functional genomics of regulation in forest trees	.64	.9	2	2.57
65	Genomics of the spruce budworm and its viral pathogens	.3	.42	1.38	1.56
66	Conifer forest health genomics	1.46	2.05	11.39	2.39

	internatio~l	INTERGE3LS	health	agriculture	environment	forestry	tech	GE3LS	BC	Prairie	ON	Quebec	Atlantic	com1	com2	com3	ABC	directed_com
34	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
35	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0
36	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
37	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1
38	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0
39	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0
40	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1
41	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1
42	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1
43	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0
44	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0
45	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1
46	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
47	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1
48	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0
49	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0
50	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1
51	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
52	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
53	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1
54	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0
55	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
56	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0
57	.	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0
58	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1
59	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0
60	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
61	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1
62	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1
63	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1
64	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0
65	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0
66	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0

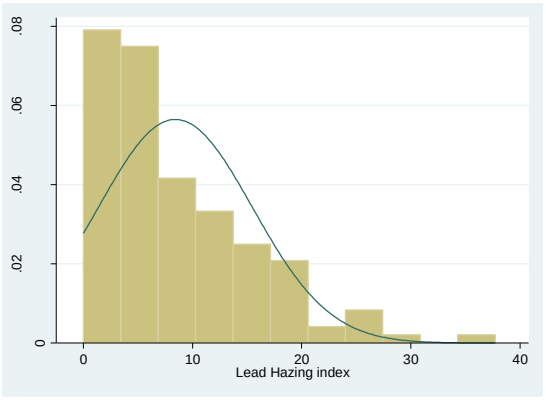
	projecttitle	GC_total	open_com	PI	research
67	Canadian program on genomics and global health	.46	.65	6.63	3.51
68	Atlantic medical genetic and genomics initiative (AMGGI)	.65	.91	6.94	1.69
69	Genomics and public health (GPH): Building public goods?	.15	.21	8.4	2.57
70	Commercialization and society and its policy and strategic implications	.24	.34	3.43	1.7
71	Building a GE3LS architecture (GE3LS Arc)	.14	.2	7.68	2.39
72	Translating science: Genomics and health systems	.19	.27	1.33	3.09
73	BEEM: Bioproducts and Enzymes from Environmental Metagenomes	.75	1.05	17.65	3.51
74	Genomics in society: Responsibilities and rights	.21	.29	7.76	2.57
75	Strengthening the role of genomics and global health	.75	1.06	6.13	3.51
76	Democracy, ethics and genomics: Consultation, deliberation and modeling	.09	.12	5.14	2.39
77	Bridging the emerging genomics divide	.21	.29	6.63	3.51
78	Segmental Duplications in neurodevelopmental, neurological and behavioral disorders	.38	.	8	3.51
79	The protist EST program	.33	.47	7.53	2.57
80	Mapping and isolation of genes influencing severity of disease in cystic fibrosis	.49	.69	10.95	3.51
81	Functional pathogenomics of mucosal immunity	1.97	2.78	1.09	1.56
82	The Canadian Pediatric Cancer Genome Consortium	.11	.	15.5	2.39
83	Atlas of genomic profiles of steroid action	1.51	2.13	19.57	2.69
84	High-throughput functional genomics using modified nucleic acid (MoNA) technologies	.41	.57	5.44	2.69
85	A quantitative and comprehensive atlas of gene expression in mouse development	.97	1.36	1.78	2.39
86	Projects in functional genomics using model organisms	.2	.28	.4	2.57
87	Genotype-specific approaches to therapy in childhood (GATC)	.62	.	15.22	2.39
88	Integrative genomics for women`s health program	.6	.84	4.52	2.69
89	Genetic determinants of human health and disease	.86	1.21	12.69	3.51
90	Montreal network for pharmaco-proteomics and structural genomics	.94	1.32	11.52	2.69
91	Functional genomics for emerging infectious diseases (Proteomics for Emerging Pathogen Response - PREPARE)	.51	.72	1.58	2.39
92	Identification of genetic pathways that regulate the survival and development of cancer and cancer stem cells	1.37	1.92	6.25	3.51
93	Structural genomics consortium (SGC phase II)	2.64	.	53	3.51
94	Public Population Project in Genomics - bridging	.07	.	24	2.69
95	Comparative and functional genomics of the human pathogen cryptococcus neoformans	.16	.22	9.1	2.39
96	A haplotype map of the human genome - biomedical tool for genetic research in Canada	1.03	1.45	3.23	2.69
97	Public Population Project in Genomics (P3G)	.1	.	24	2.69
98	Synthetic antibody program: novel therapeutics and reagents	.71	.	11.09	3.51
99	Public Population Project in Genomics (P3G phase II)	2.2	.	24	2.69

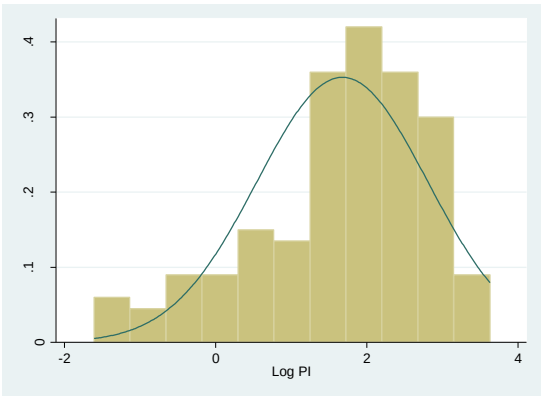
	internatio~l	INTERGE3LS	health	agriculture	environment	forestry	tech	GE3LS	BC	Prairie	ON	Quebec	Atlantic	com1	com2	com3	ABC	directed_com
67	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0
68	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0
69	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0
70	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0
71	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0
72	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0
73	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0
74	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
75	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
76	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0
77	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
78	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
79	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
80	1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
81	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
82	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
83	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
84	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
85	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
86	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
87	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
88	1	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
89	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
90	1	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
91	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
92	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
93	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
94	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
95	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
96	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
97	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
98	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
99	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1

	projecttitle	GC_total	open_com	PI	research
100	Genetic determinants of human health and disease: Annotation of chromosome 7	.26	.37	.7	3.51
101	Finding of Rare Disease Genes in Canada	.18	.	.25	1.94
102	Development and validation of comparative genomic hybridization arrays for clinical use in cancer	.34	.	8.21	2.39
103	Sequencing of the bacterium Clostridium difficile (C. difficile)	.02	.	5.36	2.69
104	Application of pharmacogenomics for rational chemotherapy of lung cancer	.5	.	.7	2.39
105	Mass spectrometer-based flow cytometer, methods and applications	.57	.	19.24	3.51
106	Efficient identification and cloning of single gene deletions in the nematode Caenorhabditis elegans	.86	1.21	6.34	2.39
107	Expression profiles of cells and tissues in C. elegans	.77	1.08	.5	.99
108	Functional annotation of essential alternatively spliced isoforms	.73	1.03	5.44	.98
109	Accelerating Genomic Innovation in Life-Science Enterprises (AGILE)	.04	.	4.17	1.72
110	Structural genomics consortium (SGC)	2.09	.	53	3.51
111	North American conditional mouse mutagenesis project: High throughput mammalian functional analysis for the discovery of novel determinants of human disease	1.23	1.73	.6	1.75
112	C. elegans: The nematode as a model organism	.31	.44	8.89	2.39
113	Functional genomics, pharmacogenomics and proteomics of the immune response in health and immune related disorders	1.02	1.43	5.19	2.69
114	Genome wide essential gene identification in candida albicans and applications to antifungal drug discovery	.39	.55	.28	2.57
115	Boosting Entrepreneurial Skills and Training: BEST in Genomics	.06	.	1.32	2.11
116	Development of Highly Active Anti-Leukemia Stem Cell Therapy Project	1.68	.	18.72	3.51
117	Pleiades promoter project: Genetic resource for CNS regional and cell specific molecular delivery	.74	1.04	2.58	2.39
118	Innovative genomic applications to develop clinical biomarkers and novel therapies for common iron metabolism disorders	.68	.	22.98	2.39
119	An integrated genetic/physical genome map for the old world monkey, cercopithecus aethiops	.2	.28	5.36	2.69
120	The stem cell genomics project	.82	1.15	15.09	1.94
121	Protein expression profiling platform for heart disease biomarker discovery	.45	.	13.57	3.51
122	NorCOMM2 - In vivo models for human disease & drug discovery	.72	.	5.61	3.51
123	Integrative biology	2.02	2.84	.2	3.51
124	The dynactome: Mapping spatio-temporal dynamic systems in humans	2	2.82	24.5	1.72
125	Functional genomics and proteomics of model organisms	1.96	2.76	14.12	3.51
126	The transplant transcriptome project	.81	.	14.19	3.09
127	High throughput mutation screening of ion channel genes in familial neurological disorders	.44	.62	12.3	2.69
128	The biomolecular interaction network database (BIND)	1.83	2.58	7.35	3.51
129	The contribution of genetic modulators of disease severity in cystic fibrosis to other diseases with similarities of clinical phenotype	.51	.72	3.85	3.51
130	Cancer genomics: A multi-disciplinary approach to large-scale high-throughput identification of genes involved in early stage cancers	1.23	1.73	4.9	2.39
131	Identifying New Genes and Medicines for the Treatment of Orphan Diseases (IGNITE)	.35	.	10.27	1.32
132	Regulatory genetics: Identification of regulatory polymorphisms in the human genome	.88	1.23	2.5	2.57

	internatio~l	INTERGE3LS	health	agriculture	environment	forestry	tech	GE3LS	BC	Prairie	ON	Quebec	Atlantic	com1	com2	com3	ABC	directed_com
100	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
101	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
102	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
103	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
104	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
105	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1
106	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
107	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
108	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
109	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
110	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
111	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
112	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
113	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
114	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
115	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
116	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
117	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
118	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
119	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
120	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
121	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1
122	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
123	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
124	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
125	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
126	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
127	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
128	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
129	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
130	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
131	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
132	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
133	Structural Genomics Consortium III														.37	.	53	3.51
134	International Regulome Consortium (IRC)														.07	.	10	1.94
135	Identification and characterization of genes involved in common developmental brain diseases														1.14	1.61	14.13	2.57
136	Functional genomics of type 1 diabetes														.8	1.12	5.08	3.51
137	Structural and functional annotation of the human genome for disease study														1.59	2.24	7.35	3.51
138	Regulatory networks in gene expression: From the genome to the organism														.78	1.11	5.13	2.57
139	Viral proteomics														.51	.71	6.78	3.51
140	Dissecting gene expression networks in mammalian organogenesis (MORGEN)														.57	.8	13.59	2.39
141	Better biomarkers of acute and chronic allograft rejection														.67	.	7.56	2.39
142	High resolution analysis of follicular lymphoma genomes														.68	.96	1.45	2.39
143	The GRID project (Gene Regulators In Disease)														.76	1.07	17.51	2.69
144	Genome-environment interactions in type 1 diabetes														1.1	1.55	5.08	3.3
145	Building the metabolomics toolbox: Enabling rapid disease diagnosis through metabolic profiling														.53	.	15.45	3.09
146	Pharmacogenomics of drug efficacy and toxicity in the treatment of cardiovascular disease														1.52	2.14	23.16	2.57
147	Therapeutic Opportunities to Target Tumor Initiating Cells in Solid Tumors														.37	.	27.92	3.51
148	Genomic tools for diagnosis and evaluation of mental retardation														.38	.	12.6	2.39
149	Autism genome project														1.14	1.6	18.31	3.51
150	International Regulome Consortium (IRC phase II)														.31	.	10	1.94
151	The pathogenomics of innate immunity (PI2)														1.18	1.65	1.09	2.39
152	Genetic dissection of complex traits using phenotypic and expression analysis of recombinant congenic mouse strains														.63	.89	8.97	2.69
153	Genomics Research Entrepreneurship to Accelerate Translation (GREAT)														.06	.	.82	2.39
154	Stratifying and Targeting Pediatric Medulloblastoma Through Genomics														.71	.	19.04	2.39
155	Proteomics and functional genomics: An intergrated approach														1.09	1.54	5.09	3.51
156	Bioinformatics of mammalian gene expression														.47	.66	5.24	2.39
133	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
134	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
135	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
136	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
137	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
138	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
139	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
140	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
141	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
142	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
143	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
144	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
145	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1
146	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
147	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
148	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
149	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
150	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
151	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
152	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
153	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
154	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
155	1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
156	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0

Appendix VI: Comparisons of PI and Log PI






```
sum  PI LGPI
```

Variable	Obs	Mean	Std. Dev.	Min	Max
PI	140	8.476	7.024197	.2	37.72
LGPI	140	1.677257	1.130392	-1.609438	3.63019

Figure 4 Comparisons of PI and Log PI

Source	SS	df	MS
Model	821.475661	1	821.475661
Residual	6036.69375	138	43.7441576
Total	6858.16941	139	49.3393483

Number of obs = 140
F(1, 138) = 18.78
Prob > F = 0.0000
R-squared = 0.1198
Adj R-squared = 0.1134
Root MSE = 6.6139

PI	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
total	4.120355	.9508178	4.33	0.000	2.240299	6.00041
_cons	5.707122	.8489492	6.72	0.000	4.028491	7.385752

Source	SS	df	MS
Model	4.67372113	1	4.67372113
Residual	172.938466	138	1.25317729
Total	177.612188	139	1.27778552

Number of obs = 140
F(1, 138) = 3.73
Prob > F = 0.0555
R-squared = 0.0263
Adj R-squared = 0.0193
Root MSE = 1.1195

LGPI	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
total	.3107912	.1609324	1.93	0.056	-.0074211	.6290034
_cons	1.468406	.1436905	10.22	0.000	1.184286	1.752525

Figure 5 Comparison of regression on PI and Log PI

Appendix VII: Regression Table

Y1: GC-TOTAL

Model A

```
. reg  GC_total  PI research international INTERGE3LS
```

Source	SS	df	MS		
Model	6.88133833	4	1.72033458	Number of obs =	155
Residual	33.3171241	150	.222114161	F(4, 150) =	7.75
				Prob > F =	0.0000
Total	40.1984624	154	.261028977	R-squared =	0.1712
				Adj R-squared =	0.1491
				Root MSE =	.47129

GC_total	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.0122717	.0040994	2.99	0.003	.0041717	.0203717
research	.0608764	.0537429	1.13	0.259	-.0453144	.1670673
internatio~l	.2326623	.0861635	2.70	0.008	.0624113	.4029132
INTERGE3LS	.26139	.0833561	3.14	0.002	.0966862	.4260938
_cons	.214284	.1456491	1.47	0.143	-.0735049	.5020729

Model B and Model C

```
. reg  GC_total  PI research international INTERGE3LS health
```

Source	SS	df	MS		
Model	10.0794349	5	2.01588698	Number of obs =	155
Residual	30.1190275	149	.202141124	F(5, 149) =	9.97
				Prob > F =	0.0000
Total	40.1984624	154	.261028977	R-squared =	0.2507
				Adj R-squared =	0.2256
				Root MSE =	.4496

GC_total	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.0109598	.0039246	2.79	0.006	.0032047	.0187149
research	.028796	.0519001	0.55	0.580	-.0737594	.1313513
internatio~l	.2004994	.082595	2.43	0.016	.0372905	.3637082
INTERGE3LS	.2526204	.0795506	3.18	0.002	.0954274	.4098135
health	.2957195	.0743467	3.98	0.000	.1488095	.4426296
_cons	.1643308	.1395127	1.18	0.241	-.1113482	.4400098

```
. reg  GC_total  PI research international INTERGE3LS health ON
```

Source	SS	df	MS		
Model	10.1501211	6	1.69168684	Number of obs =	155
Residual	30.0483414	148	.203029334	F(6, 148) =	8.33
				Prob > F =	0.0000
Total	40.1984624	154	.261028977	R-squared =	0.2525
				Adj R-squared =	0.2222
				Root MSE =	.45059

GC_total	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.0105258	.0040014	2.63	0.009	.0026185	.0184331
research	.0177327	.0552902	0.32	0.749	-.0915275	.126993
internatio~l	.2005828	.0827764	2.42	0.017	.0370065	.3641592
INTERGE3LS	.2576705	.0801833	3.21	0.002	.0992185	.4161225
health	.2946936	.0745301	3.95	0.000	.1474129	.4419742
ON	.0500469	.0848182	0.59	0.556	-.1175643	.217658
_cons	.1783783	.1418313	1.26	0.210	-.1018978	.4586543

Model D

```
. reg  GC_total PI research international INTERGE3LS health ON directed_com
```

Source	SS	df	MS		
Model	15.220553	7	2.17436472	Number of obs =	155
Residual	24.9779094	147	.169917751	F(7, 147) =	12.80
				Prob > F =	0.0000
Total	40.1984624	154	.261028977	R-squared =	0.3786
				Adj R-squared =	0.3490
				Root MSE =	.41221

GC_total	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.0184745	.0039392	4.69	0.000	.0106897	.0262593
research	.0072779	.0506173	0.14	0.886	-.0927537	.1073095
internatio~l	.2099306	.0757456	2.77	0.006	.0602396	.3596215
INTERGE3LS	.1868784	.0744899	2.51	0.013	.0396689	.3340879
health	.2594859	.0684863	3.79	0.000	.124141	.3948307
ON	.0713648	.0776922	0.92	0.360	-.0821732	.2249028
directed_com	-.4086885	.0748151	-5.46	0.000	-.5565406	-.2608364
_cons	.3169809	.132209	2.40	0.018	.0557052	.5782566

Y2: OPEN-Competition

Model A:

```
. reg  open_com PI research international INTERGE3LS
```

Source	SS	df	MS		
Model	5.20535262	4	1.30133815	Number of obs =	94
Residual	35.3787074	89	.397513567	F(4, 89) =	3.27
				Prob > F =	0.0149
Total	40.5840601	93	.436387743	R-squared =	0.1283
				Adj R-squared =	0.0891
				Root MSE =	.63049

open_com	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.0215801	.0111839	1.93	0.057	-.0006421	.0438022
research	.0656578	.0859813	0.76	0.447	-.1051852	.2365008
internatio~l	.2724842	.1505636	1.81	0.074	-.0266825	.571651
INTERGE3LS	.2634782	.1461519	1.80	0.075	-.0269225	.5538788
_cons	.5730467	.2464257	2.33	0.022	.083404	1.062689

Model B:

```
. reg  open_com PI research international INTERGE3LS  health
```

Source	SS	df	MS		
Model	9.86265579	5	1.97253116	Number of obs =	94
Residual	30.7214043	88	.349106867	F(5, 88) =	5.65
				Prob > F =	0.0001
Total	40.5840601	93	.436387743	R-squared =	0.2430
				Adj R-squared =	0.2000
				Root MSE =	.59085

open_com	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.017051	.010554	1.62	0.110	-.0039228	.0380247
research	-.0002752	.0825736	-0.00	0.997	-.1643728	.1638224
internatio~l	.2131504	.1420309	1.50	0.137	-.0691061	.4954069
INTERGE3LS	.2944623	.1372268	2.15	0.035	.0217528	.5671717
health	.4650116	.1273139	3.65	0.000	.212002	.7180211
_cons	.5203804	.2313845	2.25	0.027	.0605524	.9802085

Model C:

```
. reg  open_com PI research international INTERGE3LS  health  ON
```

Source	SS	df	MS		
Model	11.1099689	6	1.85166148	Number of obs =	94
Residual	29.4740912	87	.338782657	F(6, 87) =	5.47
				Prob > F =	0.0001
Total	40.5840601	93	.436387743	R-squared =	0.2738
				Adj R-squared =	0.2237
				Root MSE =	.58205

open_com	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PI	.0184867	.0104236	1.77	0.080	-.0022313	.0392048
research	-.0909679	.0940786	-0.97	0.336	-.2779594	.0960236
internatio~l	.2019371	.1400369	1.44	0.153	-.0764015	.4802756
INTERGE3LS	.2834623	.135304	2.10	0.039	.014531	.5523935
health	.4782359	.1256064	3.81	0.000	.2285795	.7278922
ON	.295744	.1541306	1.92	0.058	-.0106072	.6020952
_cons	.650378	.237793	2.74	0.008	.1777387	1.123017

Appendix VIII: Genome Canada Database

The detailed database about dummy variables is founded by STATA as follows:

```
. tab sector
```

Sector	Freq.	Percent	Cum.
Agriculture	16	10.26	10.26
Development of New Technologies	18	11.54	21.79
Environment	17	10.90	32.69
Fisheries	4	2.56	35.26
Forestry	11	7.05	42.31
GE3LS	11	7.05	49.36
Health	79	50.64	100.00
Total	156	100.00	

```
. tab  region
```

Region	Freq.	Percent	Cum.
Genome Alberta	9	5.77	5.77
Genome Atlantic	8	5.13	10.90
Genome British Columbia	42	26.92	37.82
Genome Prairie	12	7.69	45.51
Genome Québec	33	21.15	66.67
Ontario Genomics Institute	52	33.33	100.00
Total	156	100.00	

```
. tab  competition
```

The project is included in I, II, III, ABC or other Competition Category	Freq.	Percent	Cum.
Applied Human Health	10	6.41	6.41
Applied genomics research in Bioproduct	12	7.69	14.10
Canada/Spain Competition	3	1.92	16.03
Competition I	17	10.90	26.92
Competition II	33	21.15	48.08
Competition III	33	21.15	69.23
Entrepreneurial Education in Genomics P	3	1.92	71.15
LSP 2010 - Forestry and Environment	9	5.77	76.92
LSP 2010 - Multi-Sector	7	4.49	81.41
Not Applicable	10	6.41	87.82
Other	4	2.56	90.38
Other(Cancer Stem Cells Consortium)	2	1.28	91.67
Technology Development	13	8.33	100.00
Total	156	100.00	

Appendix IX: STATA Summary Table

```
. sum  GC_total open_com PI research international INTERGE3LS health agriculture environme
> nt forestry tech GE3LS BC Prairie ON Quebec Atlantic com1 com2 com3 ABC directed_com
```

Variable	Obs	Mean	Std. Dev.	Min	Max
GC_total	156	.6413462	.510989	.02	2.64
open_com	95	1.052211	.6634102	.12	2.84
PI	156	10.04981	9.573183	.2	53
research	156	2.511923	.739311	.43	3.51
internatio~l	155	.2967742	.458317	0	1
INTERGE3LS	156	.3205128	.4681767	0	1
health	156	.525641	.5009503	0	1
agriculture	156	.0961538	.2957516	0	1
environment	156	.1217949	.328102	0	1
forestry	156	.0705128	.2568338	0	1
tech	156	.1153846	.3205145	0	1
GE3LS	156	.0705128	.2568338	0	1
BC	156	.2564103	.4380572	0	1
Prairie	156	.1346154	.3424115	0	1
ON	156	.3397436	.4751474	0	1
Quebec	156	.2179487	.414182	0	1
Atlantic	156	.0512821	.2212828	0	1
com1	156	.1089744	.3126106	0	1
com2	156	.2115385	.4097145	0	1
com3	156	.2115385	.4097145	0	1
ABC	156	.0769231	.2673276	0	1
directed_com	156	.3910256	.4895517	0	1

REFERENCES

1 Dennis Rank, May 2009, KPMG Report: Evaluation of Genome Canada –Final Report. http://www.genomecanada.ca/medias/PDF/EN/Five-year_Evaluation.pdf

2 Genome Canada, March 2008, Genome Canada Corporate plan 2007-2008, Ottawa.

3 Genome Canada, August 2013, Performance, Audit and Evaluation Strategy 2012-2017, Ottawa.

4 KPMG LLP, March 2007, KPMG Report: Evaluation of Foundations, Prepared for Treasury Board Secretariat, Ottawa. http://www.genomecanada.ca/medias/PDF/EN/Foundations Eval - Final - Mar 14-071.pdf

5 Genome Canada, 2012, Genome Canada Corporate plan 2011-2012, Ottawa.

http://www.genomecanada.ca/medias/PDF/EN/CorporatePlan2011-12-english.pdf

6 Genome Canada, 2012, Genome Canada Annual Reports 2011-2012, Ottawa.

http://www.genomecanada.ca/medias/PDF/EN/2011-2012AnnualReport.pdf

7 Genome Canada, fall 2000, Guidelines and Evaluation Criteria for Genome Centres, Ottawa. http://www.genomecanada.ca/medias/PDF/EN/GUIDE18-final.pdf

8 Genome Canada, Dec. 2001, Guidelines and Evaluation Criteria for Competition II, Ottawa. http://www.genomecanada.ca/medias/PDF/EN/CompIIGuidelinesfinal.pdf

9 Genome Canada, 2010, Guidelines and Evaluation Criteria for Competition III, Ottawa. http://www.genomecanada.ca/data/Nouvelles/Fichiers%5Cen%5C320_1_FinalGuidelinesandEvaluationCriteria_en.pdf

10 Genome Canada, 2009, Guidelines and Evaluation Criteria - Competition in Applied Genomics Research in Bio-products or Crops, Ottawa.

http://www.genomecanada.ca/en/portfolio/research/applied.aspx

11 Phillips, P.W.B. and Eric Warren, 2010, Managing Large-Scale Science Projects: The Genome Canada Experience, Working Paper, www.VALGEN.ca.

12 Department of Finance, 2003, Budget Plan 2003, Ottawa. http://www.fin.gc.ca/budget03/bp/bpc1-eng.asp.

13 Genome Canada, June 2004, Risk Management Policy, Ottawa.

http://www.genomecanada.ca/medias/PDF/EN/RiskManagement.pdf

14 KPMG, May 2009, Evaluation of Genome Canada Final Report, Ottawa.

15 Genome Canada, Nov. 2007, Performance, Audit and Evaluation Strategy, Ottawa.

http://www.genomecanada.ca/medias/PDF/EN/PerformanceAuditandEvaluationStrategy.pdf

16 Wikipedia, "Evaluation", last modified on July 20, 2013.

http://en.wikipedia.org/wiki/Evaluation.2014.

17 Business-dictionary, "Evaluation", 2013. http://www.businessdictionary.com

18 Treasury Board of Canada Secretariat, 2014, Policy on Evaluation, http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=15024#appA

19 Arvidsson, G., 1986, 'Performance Evaluation', Guidance Control and Evaluation in the Public Sector, pp. 625–43. Berlin: Walter de Gruyter.

20 King, J. 1987, A review of bibliometric and other science indicators and their role in research evaluation, Journal of Information Science, 13, pp. 261-271.

21 Peter Henry Rossi, Howard E. Freeman, and Mark W. Lipsey, 2004, Evaluation: A Systematic Approach (7th edition), Sage Pubs ISBN: 0761908943

22 Geva-May, Iris and Leslie A. Pal, 1999, Good Fences Make Good Neighbors: Policy Evaluation and Policy Analysis - Exploring the Differences. Evaluation, 5, 3.(1999). p. 259-277

23 Birckmayer, J. and C. Weiss, 2000, Theory-Based Evaluation in Practice: What Do We Learn?, Evaluation Review (August 2000)24: 4. p.407-431.

24 Luukkonen, T., 2000, Research evaluation in Europe: State of the art. Research Evaluation (2002) 11 (2): 81-84.

25 Michael Quinn Patton, 2002, Utilization-focused Evaluation, Evaluation Models, Volume 49 (2002), pp 425-438.

26 Ruegg, R., and Jordan, G., 2007, Overview of Evaluation Methods for R&D Programs, U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Washington: DOE. http://www1.eere.energy.gov/analysis/pdfs/evaluation_methods_r_and_d.pdf

27 Jayson L. Lusk, 2005, A Meta-Analysis of Genetically Modified Food Valuation Studies. Journal of Agricultural and Resource Economics 30:1 (April 2005), p. 28-44.

28 Narongrit Sombatsompop, 2000, Research performance evaluations of Thailand national research universities during 2007-2009, Information Development (November 2010) 26:4.p. 303-313

29 Genome Canada, 2012, Genome Canada Cooperate Plan 2011-2012, Ottawa.

30 J.E. Hirsch, Sep. 2005, An index to quantify an individual's scientific research output, Physics and Society (physics.soc-ph).

31 Kahneman, D. 2002. “Maps of bounded rationality: A perspective on intuitive judgment and choice.” Nobel Prize Lecture vol. 8.