Science-Metrix

Benchmarking of Canadian Genomics - 1996–2007

Benchmarking of Canadian Genomics – 1996–2007

December 2, 2008

by

David Campbell, M.Sc.

Éric Archambault, Ph.D. and

Grégoire Côté, B.Sc.

submitted to

Genome Canada

Science-Metrix specializes in the measurement and evaluation of science, technology and innovation. Our data collection and assessment methods include bibliometrics, scientometrics, technometrics, surveys and interviews, environmental scans, monitoring and intelligence gathering. We perform program and policy evaluations, benchmarking and sector analyses, market studies and strategic planning.

514.495.6505 ▪ 1335 A Mont-Royal E. ▪ Montréal ▪ Québec ▪ Canada ▪ H2J 1Y6
info@science-metrix.com ▪ www.science-metrix.com ▪ www.rd-reports.com

Key Findings

- The scientific output in genomics rose until 1998, since which it has been stable relative to the overall production in the sciences.
- In the 2001–2007 period, Canada regained the share of the world production in genomics it had lost between 1993 to 2001.
- Canada ranked 5th in genomics for the 1996–2007 period in a multicriteria benchmarking (based on five indicators of scientific production); it ranked 6th for its number of papers, 5th for its output per capita and for its scientific impact (both observed and expected), and 10th for its specialization in the field.
- Canada ranked 6th in the most recent period (2005–2007) and will face increasing competitive pressure from a number of countries (e.g., Australia, Germany, France, Italy and China) in the future.
- Canada is in an excellent position to reap the economic benefits of genomics research and innovation: for the 12 years covered in this study, it ranked 2nd in terms of ownership of intellectual property in a multicriteria ranking (based on four indicators of technological production).
- More specifically, Canada ranked 5th for its number of patents, 3rd for its number of patents per capita and its specialization index, and 2nd for the number of citations its patents received.
- Canada took 1st place in the multicriteria ranking for its IP in genomics the most recent 3-year period (2005–2007).

Contents

Tables

Figures

1 Introduction

In the early 1950's, the discovery of the structure of DNA and the subsequent technological developments for its manipulation have had a tremendous impact on the life sciences, engendering applications from ecology to medicine. However, it is not until the modern genomics era, which began with the launch of the Human Genome Project in the late 1980's, that governing bodies started to invest substantial financial resources into genomics research (Baldy and Hatfield, 2002). These investments, together with the promise of genomics and its already substantial impact on modern science and biomedical research, have led to a number of studies comparing countries' levels of scientific and technological achievements in this area (de Looze and Ramani, 1999; de Looze *et al.*, 2001; Filliatreau *et al.* 2003; Ramani and de Looze, 2002). However, these studies focused primarily on Europe.

This report is the fifth iteration of a benchmarking study performed for Genome Canada, a not-for-profit organization supporting large-scale genomics and proteomics research projects in Canada. The initial study, performed in 1999, was based on the Thomson Reuters' Science Citation Index database (SCI). It aimed to measure, by province, the level of financial support obtained for R&D in genomics, together with the level of scientific output as measured by bibliometric methods. A second study was performed in 2000, this time aimed at examining Canada's position in genomics, as measured using published scientific papers (SCI database) and patents granted by the United States Patent and Trademark Office (USPTO). In 2003, an extensive performance monitoring scoreboard was produced. It examined the international positioning of Canada in genome science, using data extracted from Thomson Reuters' Science Citation Index Expanded database (SCIE), as well as in genomics technology, using data from the USPTO. It also drew a portrait of scientific achievements within Canada. The 2005 update reported on the salient aspects regarding Canada's positioning in genome science and technology at the world level, as part of continuing performance monitoring using the scoreboard data.

Following the announcement of the completion of the human genome sequencing in 2001, concerns regarding the ethical, economic, environmental, legal, and social implications (GE^3LS issues) of future developments in this area became more prevalent among the general public. To address these issues, Genome Canada supports large-scale GE^3LS research projects across the country and launched a GE^3LS National Program of workshops and symposiums, which has been running since 2002.

To benchmark Canada's GE^3LS research along with its core research in genomics, the current update is based on Thomson Reuters's Web of Science (WOS).[1] The Web of Science includes SCIE, which is appropriate for measuring the core of genomics research, the Social Sciences Citation Index (SSCI), and the Arts & Humanities Citation Index (A&HCI), which are appropriate for measuring GE^3LS research. The resulting benchmarking is summarized in Section 2. The benchmarking of intellectual property in genomics is again based on USPTO data, and is presented in Section 3.

---

[1] See http://www.thomsonreuters.com/products_services/scientific/Web_of_Science

2 Scientometric analysis of genomics research

This section first describes publication trends in genomics at the world and Canadian levels (section 2.1) and subsequently benchmarks leading countries in genomics and GE^3LS research (sections 2.2 and 2.3, respectively).

2.1 Publication trends in genomics

Figure 1 (left, line chart) shows that the share of genomics research in the Web of Science (WOS) increased from about 7% in 1993 to 8% in 1998, after which the scientific output in genomics, as a percentage of total scientific production, levelled out. This share decreased slightly in the most recent three-year period (2005–2007). The increase in the number of genomics papers in the WOS (Figure 1, left, column chart) reflects the expansion of the database coverage (data not shown). In 1993, there were about 43,000 genomics papers published in the WOS, compared to 73,000 in 2007.

Expressed as a proportion of the world's genomics papers (total genomics papers in the WOS), Canada's production in the field decreased slightly, from 5% in 1993 to 4.5% in 2001, after which it slowly regained its share of the world output, reaching 5% in 2007 (Figure 1, right, line chart). In 1993, Canada published about 2,100 genomics papers in the WOS, compared to 3,600 in 2007 (Figure 1, right, column chart).

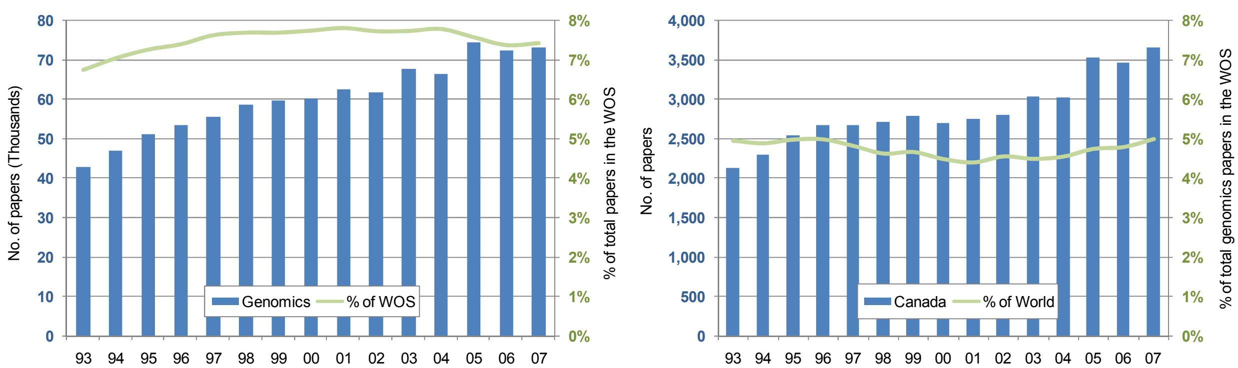

Figure 1 Genomics papers in the Web of Science (WOS) (left) and Canada's share of world papers in genomics (right), 1993–2007

Source: Compiled by Science-Metrix from the WOS

2.2 Benchmarking of scientific publications in genomics

The scientific benchmarking of the 14 leading countries is based on five indicators (see appendix for a description of the indicators, how they should be interpreted, and how they are computed):

- Number of scientific papers in genomics in the WOS – an indicator of production in the field;
- Number of papers in the WOS per million inhabitants – an indicator of productivity in the field;
- Specialization index in genomics (SI) – an indicator of the relative intensity of research in the field;

- Average of relative citations (ARC) – a measure of observed scientific impact in the field;
- Average of relative impact factors (ARIF) – a measure of expected scientific impact in the field.

Table I presents data for the 12-year period extending from 1996 to 2007, giving absolute values as well as the ranking of countries for each of the five indicators. These data indicate that Canada is well positioned in genomics. It ranks 6th in the world for number of papers, and 5th for production per capita and scientific impact (both observed and expected). The only aspect for which Canada ranks lower (10th place) is the specialization index; the proportion of its papers in genomics is about the same as the proportion of world papers in this field. In other words, although it is not specialized in genomics, Canada has a noticeable production in the field, with papers that achieved more impact than expected based on the journals in which they were published. Canada's genomics papers received, on average, 15% more citations than world papers in this field.

Table I Scientific benchmarking of leading countries, 1996–2007

Country	Papers		Papers/ 10^6 inhab.		Specialization index	ARC	ARIF	Multi rank*
United States	304,090	(1)	89	(7)	(3)	(4)	(2)	1
Switzerland	15,860	(14)	180	(1)	(4)	(1)	(1)	2
Netherlands	21,094	(11)	109	(3)	(6)	(2)	(3)	3
United Kingdom	68,941	(3)	96	(4)	(11)	(3)	(4)	3
Canada	**35,762**	***(6)***	**95**	***(5)***	***(10)***	***(5)***	***(5)***	**5**
Sweden	18,415	(12)	171	(2)	(2)	(8)	(8)	6
France	51,973	(5)	70	(8)	(5)	(9)	(6)	7
Germany	68,497	(4)	69	(9)	(7)	(7)	(7)	8
Japan	91,656	(2)	60	(10)	(1)	(12)	(12)	9
Australia	21,858	(10)	94	(6)	(12)	(6)	(9)	10
Italy	33,631	(7)	48	(11)	(8)	(10)	(11)	11
Spain	22,999	(9)	48	(12)	(13)	(11)	(10)	12
Republic of Korea	16,792	(13)	30	(13)	(9)	(13)	(13)	13
China	31,995	(8)	2	(14)	(14)	(14)	(14)	14
World	**764,819**		**10**					

Note: *Based on number of papers, papers per capita, SI, ARC, and ARIF. When the black dot is in the green area, the score is above the world average; when it is in the red area, the score is below world average (SI, ARC, and ARIF).
Source: Compiled by Science-Metrix from the WOS

Canada's positioning in genomics has been fairly stable through time with respect to each of the five indicators when the 12-year period is sliced into four 3-year periods (data not shown). More specifically, Canada lost one place for its number of papers (from 6th to 7th place) and for its specialization (from 9th to 10th place) but gained two places for its output per capita (from 6th to 4th place) and for its ARC (from 7th to 5th place) in the most recent 3-year period (2005–2007) relative to the last period of the 2005 update of this benchmarking (i.e., 2002–2004). Canada's ranking for the ARIF remained the same. Note that the observed scientific impact of Canadian papers in genomics was highest in 2005–2007.

A multicriteria benchmarking that takes into account all of the indicators presented above, in order to provide an overall assessment of the position of the leading countries in genomics over time, is presented in Table II. The US and Switzerland are the clear leaders when considering the entire period, respectively ranking 1st and 2nd overall. The Netherlands and the UK are tied for 3rd place, followed by Canada, Sweden, France and Germany (5th-8th). Although Germany is 8th overall, it improved its positioning rapidly in 1999–2001 (going from 8th to 5th, on a par with Canada) and overtook Canada in subsequent periods. In fact, Canada ranked 6th in three out of four 3-year periods. Canada, Sweden, France, and Germany are highly comparable with regard to their overall performance in genomics.

Table II Multicriteria ranking of leading countries in genomics, 1996–2007

Rank	1996-1998	1999-2001	2002-2004	2005-2007	1996-2007
1	United States	United States	United States	United States	United States
2	Switzerland	Switzerland	United Kingdom	Netherlands	Switzerland
3	United Kingdom	Netherlands	Switzerland	Switzerland	Netherlands (3)
4	Netherlands	United Kingdom	Netherlands	United Kingdom	United Kingdom (3)
5	Sweden	**Canada (5)**	Germany	Germany	**Canada**
6	**Canada**	Germany (5)	**Canada**	**Canada**	Sweden
7	France	Sweden (5)	France (7)	Sweden	France
8	Germany	France	Sweden (7)	Japan	Germany
9	Japan	Japan	Japan	France	Japan
10	Australia	Australia	Australia	Australia	Australia
11	Italy	Italy	Italy	Italy	Italy
12	Spain	Spain	Spain	Spain	Spain
13	Republic of Korea	Republic of Korea	Republic of Korea	Republic of Korea	Republic of Korea
14	China	China	China	China	China

Note: Based on number of papers, papers per capita, SI, ARC, and ARIF.
Source: Compiled by Science-Metrix from the WOS

In the 2005 update of this benchmarking, it was noted that Canada faced strong competitive pressure in genomics research; this remains the case as of 2007. Although Italy's number of published genomics papers has not yet surpassed that of Canada (they are nearly tied), it is set to overtake Canada if the observed trend continues; this is because Italy's growth, although small, is higher than Canada's growth. Moreover, Italy is actually increasing the gap which separates it from Canada in terms of specialization. Over the 12-year period, Canada has had only a slight lead on Australia in terms of papers per capita and Australia's productivity is increasing in this area. Consequently, Australia is still a serious competitor in this respect. Canada ranks highly in terms of the ARIF indicator, but France and Germany have increasingly been publishing in journals that are highly cited. In 2007, all three countries were tied for 5th place. Over time, Australian and German papers are being increasingly cited such that these countries were almost tied with Canada in terms of observed scientific impact in 2007.

China and the Republic of Korea occupy the bottom ranks of the multicriteria ranking (14th and 13th, respectively), such that they do not represent an immediate threat to Canada's global position. Still, the growth rate of their output is impressive. For example, China nearly doubled its number of genomics papers in 2005–2007 relative to 2002–2004, and it is currently the third largest producer, behind the US and Japan. China's and the Republic of Korea's positions in world science are likely to continue improving, potentially making them serious scientific competitors to Canada and to the other leading countries.

2.3 Benchmarking of scientific publications in GE^{3}LS

The scientific benchmarking of active countries in GE^{3}LS (Genomics and Ethical, Environmental, Economic, Legal and Social issues) is based on the same indicators as those used in the benchmarking of genomics research (see section 2.2). Considering the small number of papers produced in this area at the world level, only those countries with more than 30 papers are presented.

As is the case for genomics research as a whole, Canada is well positioned in GE^{3}LS research, ranking 4th on a par with Australia in the multicriteria ranking. Canada ranks 3rd for its number of papers, 5th for its output per capita and its scientific impact (both observed and expected), and 6th for its specialization. Contrary to its performance in genomics as a whole, Canada is specialized in GE^{3}LS but did not receive, on average, more citations than world papers in this area. The UK is the leader in this area, followed by the US and Denmark.

Table III Scientific benchmarking of active countries in GE^{3}LS, 1996–2007

Country	Papers		Papers/ 10^6 inhab.		Specialization index	ARC	ARIF	Multi rank*
United Kingdom	420	(2)	0.59	(1)	(1)	(2)	(2)	1
United States	1,261	(1)	0.37	(7)	(2)	(3)	(1)	2
Denmark	35	(10)	0.54	(2)	(4)	(1)	(6)	3
Australia	103	(5)	0.44	(4)	(3)	(4)	(8)	4
Canada	**156**	**(3)**	**0.41**	**(5)**	**(6)**	**(5)**	**(5)**	**4**
Netherlands	89	(7)	0.46	(3)	(5)	(8)	(3)	6
Belgium	39	(9)	0.32	(8)	(7)	(10)	(7)	7
Italy	46	(8)	0.07	(12)	(12)	(6)	(4)	8
France	96	(6)	0.13	(11)	(10)	(7)	(11)	9
Switzerland	35	(10)	0.40	(6)	(8)	(11)	(10)	9
Sweden	31	(12)	0.29	(9)	(9)	(9)	(9)	11
Germany	130	(4)	0.13	(10)	(11)	(12)	(12)	12
World	**2,509**		**0.03**					

Note: *Based on number of papers, papers per capita, SI, ARC, and ARIF. When the black dot is in the green area, the score is above the world average; when it is in the red area, the score is below world average (SI, ARC, and ARIF).

Source: Compiled by Science-Metrix from the WOS

The scientometric analysis of genomics and GE^{3}LS research reveal that Canada is an important producer of scientific output in genomics at the international level, and that its production standards are high. However, it is clear that Canada cannot afford to be complacent about its current position. It must maintain or increase its output while improving scientific excellence (as measured by scientific impact), which may prove challenging in the years ahead. Since Canada will not be able to compete with larger and rapidly developing countries in terms of quantity of output in the long run, it would be judicious for Canada to prioritize excellence.

3 Technometric analysis of genomics innovation

This section first examines trends in the number of genomics patents granted by the USPTO (section 3.1) and subsequently benchmarks leading countries based on their USPTO patents in the field of genomics (section 3.2).

3.1 Trends in patenting genomics

The number of genomics patents granted by the USPTO increased sharply from 1993 to 1998 but entered a rapid decline in 2001 (Figure 2, left, column chart). This slowdown might be explained in part by the controversy over the patenting of human genes, which culminated in 2001 with a call by the International Bioethics Committee of the UNESCO to promote the adoption of an international moratorium on the granting of human gene patents.[2] Other possible explanations include the cautious stance of governments towards patents concerning the genome, and technical problems at the USPTO, more specifically its on-going struggle to process the outstanding number of filings for patents covering expressed sequence tags from the late 1990s.[3] Only in 2006 did the trend revert, as shown by the steep increase in the number of issued patents (Figure 2). The share of genomics patents among all patents granted by the USTPO followed a similar trend. In 2007, this share was back near its 1998 peak of about 3% (Figure 2, left, line chart).

Canada's number of genomics patents granted by the USPTO followed the same general trend (Figure 2, right, column chart). While Canada owned 2.2% of genomics patents issued by the USPTO in 1993, it owned 3.3% of these patents in 2007 (Figure 2, right, line chart).

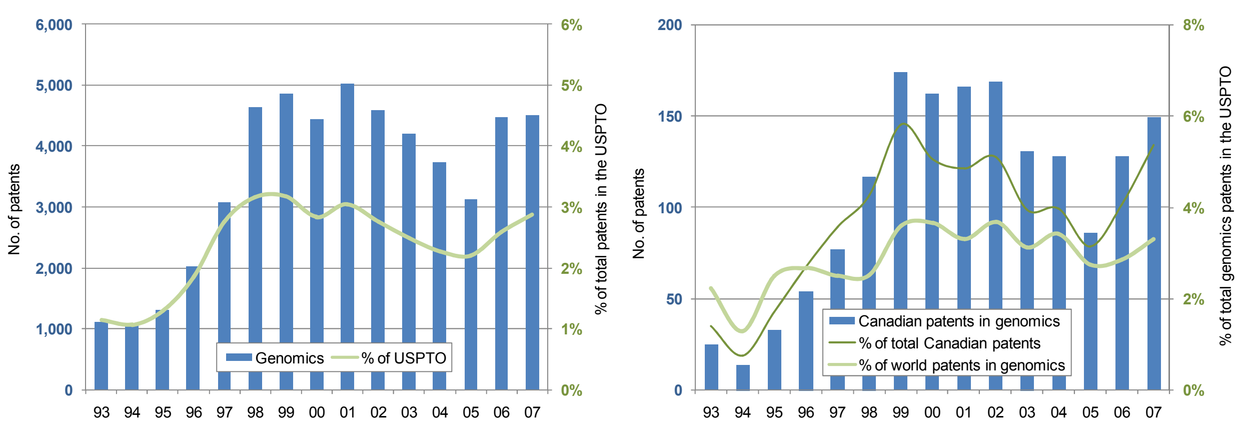

Figure 2 Number of patents granted by the USPTO in genomics (left) and Canada's share of genomics patents (right), 1993–2007

Source: Compiled by Science-Metrix from USPTO data

---

[2] See http://www.gene.ch/genet/2002/Jan/msg00020.html

[3] See http://www.biotech-info.net/geno-types.html

3.2 Benchmarking of intellectual property in genomics

A method similar to that which was used to benchmark Canada's scientific output in genomics is applied here in order to benchmark genomics technology. The ranking of countries with respect to their number of granted patents, their number of patents per capita, their specialization index, and their average of relative citations is analysed over a 12-year period (Table IV). These indicators are subsequently combined in a multicriteria benchmarking to appreciate the overall performance of countries over time (Table V). It should be noted that although the USPTO database presents an obvious bias towards the US, it is a potent tool for comparing other countries with one another since the most important inventions are usually patented to obtain protection in the largest market in the world (i.e., the US market).

Table IV demonstrates that Canada has been an important player in the acquisition of intellectual property rights on genomics inventions between 1996 and 2007. Indeed, together with Japan, Germany, the UK, and France, Canada is one of the foremost producers of genomics-related intellectual property (IP), as measured by patents granted by the USPTO.

Table IV Technological benchmarking of leading countries, 1996–2007

Country	Patents		Patents/ 10^6 inhab.		Specialization index	ARC	Multi rank*
United States	34,756	*(1)*	10.1	*(1)*	*(6)*	*(1)*	1
Canada	**1,541**	***(5)***	**4.1**	***(3)***	***(3)***	***(2)***	**2**
United Kingdom	1,580	*(4)*	2.2	*(6)*	*(1)*	*(3)*	3
Netherlands	643	*(7)*	3.3	*(5)*	*(5)*	*(5)*	4
France	1,511	*(6)*	2.0	*(9)*	*(4)*	*(7)*	5
Japan	3,317	*(2)*	2.2	*(7)*	*(11)*	*(6)*	5
Sweden	384	*(10)*	3.6	*(4)*	*(8)*	*(4)*	5
Australia	479	*(9)*	2.1	*(8)*	*(2)*	*(8)*	8
Switzerland	499	*(8)*	5.7	*(2)*	*(7)*	*(10)*	8
Germany	2,001	*(3)*	2.0	*(10)*	*(9)*	*(9)*	10
Republic of Korea	356	*(11)*	0.6	*(11)*	*(12)*	*(11)*	11
Italy	204	*(12)*	0.3	*(12)*	*(10)*	*(12)*	12
World	**48,668**		**0.7**				

Note: *Based on number of patents, patents per capita, SI, and ARC. When the black dot is in the green area, the score is above the world average; when it is in the red area, the score is below world average (SI and ARC).
Source: Compiled by Science-Metrix from USPTO data

Canada is specialized in the technological aspect of genomics; its proportion of USPTO patents in this field being 65% higher than the overall proportion of genomics patents in the USPTO database (Table IV). During the 12-year period, Canada ranked 3rd for its number of patents per capita and 2nd for its average of relative citations (ARC). Overall, Canada has proven to be a strong performer, ranking 2nd behind the US in the multicriteria benchmarking.

When the 12-year period is sliced into four 3-year periods, Canada's performance has been strongest in the most recent three-year period (2005-2007). In fact, Canada joined the US in 1st place in this period as a result of an increase in its number of patents and its ARC (Table V). More precisely, Canada's patents received 25% more citations, on average, than genomics patents in the USPTO in 2005–2007. Over the 12-year period, Australia has consistently risen in the ranking, reaching the 6th place in 2005–2007. There is no doubt that Australia could become a serious competitor for Canada in the coming years, whereas the UK stands to remain in the lead group having had a stable performance through time. Thus, although Canada is currently well-positioned for its IP in genomics, considerable effort will be required if it is to stay in the lead.

Table V Multicriteria ranking of leading countries for their IP in genomics, 1996–2007

Rank	1996-1998	1999-2001	2002-2004	2005-2007	1996-2007
1	United States	United States	United States	**Canada (1)**	United States
2	Netherlands	**Canada (2)**	**Canada**	United States (1)	**Canada**
3	**Canada**	United Kingdom (2)	United Kingdom	Switzerland	United Kingdom
4	United Kingdom	Netherlands	France	Japan (4)	Netherlands
5	Switzerland	France (5)	Netherlands	United Kingdom (4)	France (5)
6	Sweden	Sweden (5)	Switzerland	Australia	Japan (5)
7	France (7)	Switzerland	Germany	France (7)	Sweden (5)
8	Japan (7)	Australia	Australia (8)	Germany (7)	Australia (8)
9	Australia (9)	Germany	Japan (8)	Netherlands (7)	Switzerland (8)
10	Germany (9)	Japan	Sweden	Sweden (7)	Germany
11	Italy	Republic of Korea	Republic of Korea	Republic of Korea	Republic of Korea
12	Republic of Korea	Italy	Italy	Italy	Italy

Note: Based on number of patents, patents per capita, SI, and ARC.
Source: Compiled by Science-Metrix from USPTO data

4 Conclusion

Since the 2005 update of this benchmarking—when scientific activity in genomics had been stable for about seven years—the relative prominence of this field within the sciences as a whole has decreased slightly. However, this is too faint a change over too short a period to conclude that the stability of the scientific production in this area is compromised. While Canada's share of the scientific output in genomics decreased from 1993 to 2001, it gradually regained its lost share of world output from 2001 to 2007.

The data presented in this scoreboard show that Canada is an important producer of scientific knowledge in genomics at the international level, and that it has been a solid producer of first-rate science in the last decade. It is ranked 5th behind the US, Switzerland, the Netherlands and the UK in a multicriteria ranking, and it performs best in terms of number of published papers (6th), output per capita (5th) and scientific impact (5th). However, Canada is facing increasing pressure from other countries that have increased their production, productivity, specialization and/or scientific impact in the last 12 years. Among them, Australia, Germany, France and Italy are worth mentioning. In the 2005 update, it was recommended that Canada should prioritize quality rather than quantity in order to maintain its competitiveness internationally, as it was shown that some emerging countries were set to overtake many of the leaders in terms of production volume and that Canada would not be able to compete with them in the long run. This conclusion is now more solid than ever, with China having nearly doubled its yearly scientific production in 2007 compared to 2004 and having overtaken most of the leaders including Canada. China actually ranked 3rd for its number of papers in 2007, following close behind Japan. It therefore seems that it will be increasingly challenging for Canada to maintain its position in the future.

In terms of ownership of intellectual property rights in genomics, Canada ranked 2nd for the 1996–2007 period and 1st in the most recent 3-year period (2005–2007). This positioning of Canada reflects its very good performance across the whole set of technometric indicators. Thus, in addition to setting and adhering to high standards for scientific research, Canada also has a strong IP position, which provides a robust base for profiting from the commercialization of the results of genomics research.

References

Baldy, P. and Hatfield, G.S. 2002. DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling. Cambridge University Press, 207 pages.

de Looze, M.-A. and Ramani, S.V. 1999. Biotechnology patent application in Europe – A look at the differences between French, British, and German patent application trends. *Nature Biotechnology.* 17(1): 83-85.

de Looze, M.-A., Coronini, R. and Joly, P.-B. 2001. A note on recent trends in knowledge creation and appropriation thought genomics: a scientometric analysis. *International Journal of Biotechnology.* 3(1-2): 457-480.

Filliatreau G. et al. 2003. *Bibliometric Analysis of Research in Genomics during the 1990's*, Study produced by OST in collaboration with the LTSI, at the request of the CNRG and the RNG, 53 pages.

Ramani, S.V. and de Looze, M.-A. 2002. Country-specific characteristics of patent applications in France, Germany and the UK in the biotechnology sector. *Technology Analysis and Strategic Management.* 14(4): 457-480.

Appendix - Methods

For the benchmarking of countries in genomics research, the decision was made to select the 14 countries that were most active in genomics during the last 12 years. Earlier studies by Science-Metrix have included the Russian Federation and Belgium. However, they are no longer in the top 14, having been overtaken by China and the Republic of Korea. As for genomics technology, China and Spain were not included in the multicriteria benchmarking because they did not obtain a sufficient number of patents in each of the three-year periods analysed.

Keywords used to build the datasets

The datasets were constructed by queries, using genome-specific keywords in paper titles for the science dataset, and keyword-in-title and keyword-in-abstract searches for the patent dataset. The keyword set for the query was originally defined in 1999 by experts appointed by Genome Canada and by Science-Metrix analysts. This keyword set was extensively revised (in 2000 and in 2003) to remove as many false positives as possible and to achieve maximum coverage. The resulting datasets comprise papers and patents in core and peripheral genomics (such as papers in molecular biology that touch upon genome research). Some keywords were used only within the boundary of genomics (e.g. the term "sequenc*" can apply, for instance, to mathematics; therefore these papers were excluded). The resulting keyword set is listed below (the * represents a wildcard character, which means that any word starting with the letters preceding the * will be included).

2-hybrid	Genom*	Mutation	Proteom*
Allel*	Genot*	Nucleic acid	QTL
Antisense	Haplotyp*	Nucleosid	Radiation hybrid map
Autosom*	Intron	Nucleotid*	Recombinant
Biochip	Linkage map	PCR	Restriction map
Candidate region	Loci	Physical map	RNA (mRNA, RNAi, etc.)
Chromoso*	Locus	Plasmid*	Sequenc*
Cloning	Meiosis	Ploidy	SNP
DNA (cDNA, anti-DNA, etc.)	Meiotic	PNA	Transcript*
Exon	Microsatellite	Polymerase	Transgen*
Expressed	Minisatellite	Post-translation	YAC
Genetic*	Molecular characterization		

The GE[3]LS dataset on the ethical, economic, environmental, legal, and social issues raised by genomics research is a subset of the previously defined genomics dataset. Relevant genomics papers were selected by using keyword-in-title searches and by scanning genomics papers classified within the humanities and social sciences based on the journal classification used by the US National Science Foundation (NSF) in producing its Science and Engineering Indicators series (see: http://www.nsf.gov/statistics/seind06/).

Data and methods used for the scientometric analysis

This scientometric analysis is based on the use of Thomson Reuters's Web of Science (WOS).[4] The Web of Science includes the Science Citation Index Expanded (SCIE), which is appropriate for

[4] See http://www.thomsonreuters.com/products_services/scientific/Web_of_Science

measuring the core of genomics research, the Social Sciences Citation Index (SSCI), and the Arts & Humanities Citation Index (A&HCI), which are appropriate for measuring GE^3LS research. The WOS indexes some 8,700 journals. These journals are considered to be the most important peer-reviewed journals in their respective fields. The statistics are drawn from three types of documents considered to be original contributions to scientific knowledge: articles, notes, and reviews. The tables presented herein refer to these three types of documents as "papers". The dataset was used to produce detailed statistics based on the following indicators:

Number of papers (Papers) - Number of scientific papers written by authors located in a given geographical, geopolitical, or organizational entity (e.g., a country or an institution).

Specialization index (SI) - An indicator of the intensity of research in a given geographic or organizational entity relative to the overall output in the reference. For example, if the percentage of Canadian papers in the field of genomics is greater than the percentage of papers in this field at the world level (the reference), then Canada is said to be specialized in this field.

Average of relative impact factors (ARIF) - This indicator is a proxy for the expected impact of an entity's papers based on the journal in which they are published. For each journal, an impact factor (IF) is calculated based on the number of citations it received relative to the number of papers it published. Thus, each journal's IF varies from year to year. The IF of papers is calculated by ascribing to each of them the IF of the journal in which they were published, for the year in which they were published. To account for different citation patterns across fields and subfields of science (e.g., there are more citations in biomedical research than mathematics), each paper's IF is divided by the average IF of the papers in its subfield to obtain the Relative Impact Factor (RIF). The ARIF of a given entity is the average of the RIFs of the papers belonging to it. When the ARIF is above 1, it means that an entity scores better than the world average; when it is below 1, it means that on average, an entity publishes in journals that are not cited as often as the world level.

Average of relative citations (ARC) - The ARC is an indicator of the scientific impact of papers produced by a given entity (e.g., a country or an institution). The number of citations received by each paper is counted for the year in which it was published and for the two subsequent years. For papers published in 1996, for example, citations received in 1996, 1997 and 1998 are counted. The only exceptions are 2006, which has a citation window of two years (2006 and 2007), and 2007, which has a citation window of one year, because citation data are not yet available for the subsequent years. To account for different citation patterns across fields and subfields of science (e.g., there are more citations in biomedical research than in mathematics), each paper's citation count is divided by the average citation count of the papers in its subfield to obtain a relative citation count (RC). The ARC of a given entity is the average of the RCs of the papers belonging to it. An ARC value above 1 means that a given entity is cited more frequently than the world average, while a value below 1 means the reverse.

Data and methods used in the technometric analysis

Patents are often used as a measure of invention. However, they are known to have a number of limitations, including:

- incompleteness: many any inventions are not patented, since patenting is only one way of protecting an invention;
- inconsistency in quality: the importance and value of patented inventions vary considerably;

- inconsistency across industries and fields: industries and fields vary considerably in their propensity to patent;
- inconsistency across countries: inventors from different countries have different propensities to patent inventions, and different countries have different patent laws.

Despite these limits, patents are widely used to compare the level of technological development of different geographic and organizational entities. This report uses the US Patents and Trademark Office (USPTO) database. Its data are commonly used to measure invention, since the USPTO has one of the largest registers of patented inventions in the world. Because the US is the largest market in the world, the most important inventions tend to be patented there. Although the USPTO database presents an obvious bias towards the US, it is still a potent tool for country-level comparison.

The statistics presented here concern utility patents that have been granted (and not patent applications). The dataset was used to produce detailed statistics based on the following indicators:

Number of patents (IP) - Unlike scientific publications, patents possess two fields that contain bibliographic information relevant to the calculation of where a patent originates: the inventor field and the assignee field. These fields can be used to compute statistics on two different indicators—namely, invention and intellectual property (IP). The majority of patents are owned by corporations, and their addresses, which appear in the assignee field, are used to compute the geographical location of the ownership of IP. In some cases, where an individual owns the IP, the address of this owner is used to compute the location of the IP. For the sake of simplicity, this report presents data on IP only.

Specialization index (SI) - This is an indicator of the concentration of IP in a given geographic or organizational entity relative to the overall IP for a given reference. For example, if the percentage of Canadian patents in the field of genomics is greater than the percentage of patents in this field at the world level (the reference), then Canada is said to be specialized in this field.

Average of Relative Citations (ARC) – The number of citations received for each patent were counted for the year in which they were granted and the two years that followed. For patents granted in 1996, for example, citations received in 1996, 1997, and 1998 were counted. The only exception is the year 2006, which contains a citation window of two years, and 2007, which contains a citation window of one year. The number of citations of each patent was normalized by the average number of citations of patents issued the same year to obtain a relative citation count (RC). The ARC of a given entity is the average of the RCs of the patents belonging to it. An ARC value above 1 means that a given entity is cited more frequently than the world average, while a value below 1 means the reverse.

For both papers and patents, data that are weighted per capita at the country level are based on population statistics produced by the US Census Bureau. These statistics present annual data for every country, estimated at mid-year.