Characterization of Genomics in Canada

A Bibliometric Study of
Scientific Articles and Research Grants
1995-1997

Éric Archambault, Yves Gingras,
Benoît Godin, François Vallières

Produced for Genome Canada

August 1999

Observatoire des sciences et des technologies
3465, rue Durocher Montréal (Québec) H2X 2C6
Téléphone : (514) 499-4074 Télécopieur : (514) 499-4065
www.ost.qc.ca

This document as well as other recent publications by OST can be obtained from the following internet site: http://www.ost.qc.ca

Table of Contents

Tables

Executive summary

This bibliometric study shows that genomics in Canada is characterized by a high level of concentration. Most of the research is undertaken and it is mostly used in the health sector. The university and the hospital sectors contribute to a large part of the production while the private sector is not highly active in the field. In terms of technological platform, the field of genomics in Canada is highly developed in functional genomics but may not be sufficiently developed in enabling technologies (that is, bioinformatics and new technologies). Forestry, fisheries and agriculture do not account for an important share of research in genomics and this is somewhat surprising considering the economic importance of these sectors to the Canadian economy.

The leading provinces are by far Quebec and Ontario whilst Alberta and British Columbia follow at a distance. Quebec seems to provide a fertile environment that combines strong university research with higher than average involvement from the hospital and the private sectors. Many of these indicators converge to suggest that Quebec specializes more in genomics than other provinces although Ontario still leads in terms of absolute numbers in several areas. Alberta is consistently performing well suggesting it ranks third in genomics. British Columbia receives more grants but is not as productive as one could expect from this important input. Other provinces do not seem to have reached a critical mass whereby research in genomics could become self-sustaining and provide world-class results in important quantities.

The Observatoire des sciences et des technologies has been mandated by Genome Canada to perform a bibliometric study of the field of genomics in Canada. The first part of this report describes the method used to constitute the datasets, how the statistics were computed and mentions some caveats associated with the study. The second part presents the salient characteristics of research in genomics at the provincial level.

1.0 Methodological notes

Section 1.1 describes the method used to constitute the datasets while Section 1.2 tells how the statistics were computed.

1.1 METHOD USED TO CREATE THE DATASETS

The method used to constitute the datasets for this study involved searching the titles of scientific articles or grants contained in relevant databases. We undertook keyword searches in five databases:

1. Science Citation Index (SCI)
2. Social Science Citation Index (SSCI)
3. Arts and Humanities Citation Index (AHCI)
4. Natural Science and Engineering Research Council (NSERC) database of grants
5. Medical Research Council (MRC) database of grants in the health sector

Three datasets were constructed, leaving both the AHCI and the SSCI datasets aside. The rejection of the AHCI and the SSCI data was motivated by two reasons. Firstly, those articles were not necessarily exclusive to those databases. For example, the American Journal of Medical Genetics as well as the journal Behavior Genetics could also be found in the SCI database. Secondly, the number of articles were fairly low and therefore precluded the production of robust statistics, particularly for disaggregated data – there were only 42 articles including the keywords in title in the AHCI and 329 in the SSCI.

The SCI database was used during the definition stage of the project. During that stage, the *Observatoire* worked hand-in-hand with Genome Canada's Scientific Committee in the definition of a set of keywords that would represent the field of genomics as precisely as possible. As a starting point, the *Observatoire* visited the Web site of *The Institute for Genomic Research* (TIGR – Stanford University, California) and defined a list of six keywords that would allow the selection of about 90% of the articles mentioned in the list of publications of this institute (www.tigr.org/about/publications.html). These keywords appear in 57% of the valid articles in the resulting SCI dataset[1]. With the help of the Scientific Committee, a list of

[1] These words were: genom*, gene(s), sequenc*, receptor*, *DNA, *RNA. The complete list of keywords used to constitute the datasets is annexed to this document.

more than 50 keywords was added to the initial list of 6. Not all these terms were added at the same time and we could thus see that diminishing returns were reached, that is, adding new keywords did not translate in many new items. Hence, the resulting list of keywords has the capacity to find most of the articles in genomics that did not use esoteric titles.

The main difficulty involved finding keywords that were specific to the field of genomics - a difficult task because genomics is a multidisciplinary field and as such it uses the same terms as, for example, the field of molecular biology, which comprises research in genomics but also activities outside that field. Another difficulty is that it is hardly possible to find all relevant articles using only terms that are specific to the field of genomics. For example, the term "sequence" is used in genomics but also in geology. Similarly, the term "characterization" is used often in genomics but is also used in chemistry, in electronics and in physics. In particular, the term "protein" was deemed important because proteomics is a branch of genomics. However, this term is not used specifically and it was not possible to define a method to extract only those articles containing the term protein in their title that were specific to the field of genomics. Thus, it was decided to drop that keyword even though it also meant not selecting articles in proteomics.

Because the datasets comprised records outside the field of genomics, we had to undergo a stage of manual cleaning. This comprised essentially two steps: going through records one by one to eliminate those that appeared to be outside the field of genomics; searching for items containing specific keywords which appeared problematic (e.g. sequence; characterization; epigenetic) and verifying the validity of those records once again. This procedure was associated with a substantial reduction of the number of valid records in the datasets (Table 1).

Table 1 Number of items in the datasets

Database	Items in initial dataset	Items in resulting dataset	Percentage removed
NSERC	1 113 grants	658 grants	42%
MRC	2 888 grants	1 586 grants	45%
SCI	16 487 papers	12 081 papers	27%

Source: Observatoire des sciences et des technologies

Despite the care taken in the construction of the datasets, there are limits inherent to this method:

A) the datasets still contain articles or grants that do not belong to the field of genomics per se

B) the datasets omit articles or applications for grants in the field of genomics

i) that did not use the selected set of keywords in their title

ii) that did not appear in journals indexed in the SCI; or that were not submitted to organisations whose grants are indexed in the NSERC or MRC database

In order to characterize the data and to produce statistics on the proportion of papers in different platforms and sectors of activity, we sent a random sample of 500 articles to a panel of five experts in genomics. Depending on how it is computed, the first type of limit is associated with a 15-24% overestimation in the case of the SCI dataset[2]. The very high level of variation in the answers of the different experts to the question of whether these articles belong to genomics shows just how difficult it is to define the boundaries of this emerging scientific field (Table 2).

Table 2 Number of papers in genomics according to five experts

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Papers related to genomics	426	377	418	225	449
Unrelated papers	74	123	82	275	51
Overestimation	15%	25%	16%	55%	10%

Source: Observatoire des sciences et des technologies

Currently, we do not have a way of measuring the underestimation associated with the second type of limit. Nevertheless, this underestimation should not be too important because : 1) the list of keywords used is exhaustive and adding new terms to the search expression increased only marginally the number of articles found; 2) most of the important journals are covered by the SCI and most of the Canadian researchers in genomics who apply for grants do so with the NSERC or one of the organisations contained in the MRC database[3].

[2] Averaging the score of these experts returns an overestimation of around 24%. When two outliers are removed (experts 4 and 5), the overestimation falls down to less than 19%. On the other hand, if we look directly at the dataset and ask how many items are considered as unrelated to genomics by 3 out of 5 experts, the overestimation falls down to 17%. Excluding the score of two experts who diverge most from the others leads to an overestimation of around 15%, that is, when 2 out of 3 experts claim that an article is irrelevant.

[3] These organisations are: Alberta Cancer Board; Alberta Cancer Foundation; Alberta Heritage Foundation for Medical Research; Alzheimer Society of Canada; Amyotrophic Lateral Sclerosis Society of Canada; Arthritis Society; Atkinson Charitable Foundation; B.C. Health Research Foundation; Banting Research Foundation; Bayer/Canadian Red Cross Society Research and Development Fund; Canadian Cystic Fibrosis Foundation;

These limits, although not to be dismissed, should not adversely affect the broad results of this study. For example, the SCI dataset comprises more than 12 000 articles. This is enough to characterize the efforts of all the provinces in Canada as well as those of the largest cities and of the most productive institutions and authors in the field. Further, relying on relative measures of the field (percentages) as opposed to giving too much credence to absolute figures, the statistics thus computed remain fairly robust since, *ceteris paribus*, these limits affect provinces, institutions, cities and authors in a similar manner. More care should be given to the interpretation of absolute figures, particularly those covering subsidies in the field.

1.2 METHOD USED TO COMPUTE THE STATISTICS

The computation of statistics from the random sample involved calculating an arithmetic average of the answers from the panel of experts in order to obtain the proportion of articles produced on each technological platform and within each sector of activity. One must be careful when interpreting these statistics because there were important variations in the answers from the experts. These variations are compounded by the fact that relying on five experts means that each one weighs strongly on the average score. Table 3 shows the effect of removing one or two experts from the dataset on the number of items where agreement was reached. For example, when all the experts are included, there are only 47 articles where they agreed on the technological platform. Removing two outliers, the number of articles where an agreement was reached on the platforms jumps to 160.

Canadian Diabetes Association; Canadian Liver Foundation; Canadian Lung Association; Canadian Nurses Foundation; Canadian Psychiatric Research Foundation; Canadian Red Cross Society; Cancer Research Society Inc.; Crohn's and Colitis Foundation of Canada; Dairy Farmers of Canada; Dalhousie Medical Research Foundation; EJLB Foundation; Fonds de la recherche en santé du Québec; Fonds pour la formation de chercheurs et l'aide à la recherche; Health Canada-National Health Research & Development Program (NHRDP); Health Services Utilization and Research Commission; Heart and Stroke Foundation of Canada; Hospital for Sick Children Foundation; Huntington Society of Canada; J.P. Bickell Foundation; Juvenile Diabetes Foundation Canada; Kidney Foundation of Canada; Manitoba Health Research Council; Manitoba Medical Service Foundation; Max Bell Foundation; Medical Research Council of Canada; Multiple Sclerosis Society of Canada; Muscular Dystrophy Association of Canada; National Cancer Institute of Canada; Ontario Mental Health Foundation; Ontario Ministry of Health; Parkinson Foundation of Canada; Physicians' Services Incorporated Foundation; PMAC - Health Research Foundation; Rick Hansen Man in Motion Foundation; RP Eye Research Foundation; St. Boniface General Hospital Research Foundation Inc.; Vancouver Foundation: British Columbia Medical Services Foundation; Vancouver Foundation: W.J. VanDusen Foundation.

Table 3 Effect of removing outliers on experts' agreement on three questions

Experts	Relevance	Platform	Sector	All three questions
1,2,3,4,5	121	47	100	17
1,2,3,5	213	123	177	52
1,2,3	301	160	206	91

Source: Observatoire des sciences et des technologies

Other statistics used the three complete datasets and as such possess both type A and type B limits. In order to lower the effect of these limits, two methods were employed:

using percentages instead of absolute figures at the provincial aggregation level;

when listing cities, institutions, and researchers, retaining only those which appear several times.

The effect produced on those listings depends on the nature of the variable being observed. At the city and institution level, the effect should be spread relatively evenly across the board. In contrast, there is less chance that authors outside the field appear in the list of publications because the numbers are large and, to be included, a researcher had to publish at least 9 articles (that is, the equivalent of 3 per year on average). Since these authors are seemingly from the field of genomics, there is not a high probability that the list will comprise articles being outside of genomics. The probability of including researchers outside the field of genomics is greater in the case of grants because we have not excluded those who have only one grant in the field and therefore desegregated statistics on grants should be interpreted with more care. Due to the large numbers involved, we can make stronger claims with the statistics drawn from scientific articles.

2.0 Salient characteristics of genomics in Canada

This second part of the report describes the salient characteristics of the field of genomics that we can infer from the data gathered in this bibliometric study. We will firstly look at a number of figures representing inputs in the field of genomics (subsidies, researchers) and, secondly, at measures of output, namely Canadian articles in genomics indexed in the SCI database.

Section 2.1 examines the data on grants in the field of genomics by province. Section 2.2 examines the number of times that articles by researchers who have authored at least 9 articles mention Canadian institutions. This section examines the number of leading researchers by province and compares these data to the population as well as the level of subsidy.

Section 2.3 examines in which sectors and in which provinces these scientific articles were produced and compares this output to the whole of Canadian production as indexed in the SCI database for 1995-1997. Finally, Section 2.4 examines how scientific articles in genomics are distributed by technological platforms and by sector of use.

2.1 GRANTS IN THE FIELD OF GENOMICS

Because grants are obtained after a peer review process, they provide a potent indicator of the state of a field, more particularly when these data are aggregated at the provincial level. The data from the Medical Research Council database is particularly useful because it comprises the grants of a relatively large set of institutions, each with their own different agenda. Combined with data from the NSERC, these grant data provide a robust picture of the strength of genomics in the different provinces.

In terms of grants indexed in the MRC database, Ontario comes first in absolute number although being almost neck-to-neck with Quebec, each province receiving about $45 million in grants for research in genomics. British Columbia comes third with more than $25 million in grants from health authorities while Alberta remains in the race with close to $12 million received over three years. Other provinces receive less than 4% each. Not surprisingly, smaller provinces receive smaller grants.

The data from the NSERC shows that Ontario leads over Quebec by a wider margin, with more than $20 million against $15 million in the case of Quebec. British Columbia obtains nearly $9 million while Alberta obtains close to $6 million. When the grants from the NSERC and those in the MRC database are combined, these tendencies are preserved. In absolute number, Ontario leads the way, followed by Quebec, British Columbia and Alberta. Manitoba, Nova Scotia and Saskatchewan follow at a distance while Newfoundland, New Brunswick and Prince Edward Island each receive less than $1 million.

These data can be weighed against the population of each province. Particularly worthy of attention is the ratio of the percentage of grants by province against the percentage of the Canadian population. The greater the ratio, the greater the grants in genomics per capita. Table 4 clearly shows that the leader in this area is British Columbia with 1,4 times the amount of grants relative to its population within Canada. This is followed by Quebec with 1,2 times the grants relative to its proportion of Canadian population. Alberta, Nova Scotia and Ontario obtain a level of grants in genomics, which is consistent with the proportion of their population within Canada. Saskatchewan, Newfoundland, Prince Edward Island and particularly New Brunswick receive a smaller share of Canadian grants in genomics in relation to their share of the Canadian population.

Table 4 Ratio of grants (1995-1997) to population (1996) by Province

Province	% Population	% Grants	%Grants/ %Population
Alberta	9,4%	9,0%	1,0
British Columbia	13,1%	17,9%	1,4
Manitoba	3,8%	3,3%	0,9
New Brunswick	2,5%	0,1%	0,0
Newfoundland	1,9%	0,4%	0,2
Nova Scotia	3,1%	3,2%	1,0
Ontario	37,4%	34,3%	0,9
Prince Edward Island	0,5%	0,1%	0,2
Québec	24,5%	29,9%	1,2
Saskatchewan	3,4%	1,9%	0,5
TOTAL	100%	100%	1,0

Source : Observatoire des sciences et des technologies

Grant data are interesting because they provide the opinion of peers relative to the worth of research projects whilst simultaneously providing a direct indicator of input in a field. The data show that Ontario and Quebec are the clear leaders in genomics in Canada. Following at a distance are British Columbia, which fares well relative to its share of Canadian population. Alberta is also in the race whilst Manitoba and Saskatchewan are trailing. The smaller provinces cannot be seen as taking part in the race

2.2 CANADIAN RESEARCHERS IN GENOMICS

The number of researchers in each province is another measure of input that can be used to evaluate the field of genomics. These statistics comprise the 650 most active Canadian authors in the field of genomics as indexed by the SCI in 1995-1997. Together, these researchers who represent only one percent of the authors who gravitate in the Canadian field of genomics research (including collaborators from other countries) but have authored around two thirds of the papers in the SCI dataset of more than 12 000 articles.

In terms of absolute as well as relative figures, Quebec is the province with the largest share of researchers (Table 5). Alberta and British Columbia follow Quebec and Ontario at a distance in terms of absolute figures. In terms of researchers by the share of Canadian population, Quebec clearly lead the way with 1,6 times as many researchers as its share of the Canadian population. Alberta and Ontario have a share of active researchers that is consistent with the share of their population within Canada while British Columbia, Manitoba, Saskatchewan and the smaller provinces do not have a share of active researchers that follows their share in terms of population.

When it comes to the ratio of the percentage of grants to the percentage of active researchers, British Columbia and Saskatchewan clearly receive more grants than their percentage of researchers. In the case of Nova Scotia, the numbers involved are really low and one cannot make any strong claim about this ratio. Alberta, Ontario and more particularly Quebec do not receive as high a percentage of grant as their percentage of the active researchers.

This evidence suggests that only Alberta, British Columbia, Ontario and Quebec may have a critical mass of active researchers in genomics. According to this indicator, Quebec leads the way in terms of active researchers in genomics.

Table 5 Ratio of researchers (1995-1997) to population (1996) by Province

Province	Researchers 9+ articles	% Researchers 9+ articles	%Researchers / %Population	%Grants / %Researchers
Alberta	62	9,7%	1,0	0,9
British Columbia	54	8,4%	0,6	2,1
Manitoba	13	2,0%	0,5	1,6
New Brunswick				
Newfoundland	3	0,5%	0,2	0,8
Nova Scotia	8	1,2%	0,4	2,6
Ontario	243	37,9%	1,0	0,9
Prince Edward Island	1	0,2%	0,3	0,5
Québec	251	39,1%	1,6	0,8
Saskatchewan	7	1,1%	0,3	1,7
TOTAL	642	100%	1,0	1,0

Source : Observatoire des sciences et des technologies

2.3 GENOMICS BY SECTORS OF PRODUCTION

It is interesting to note that most of genomic research is carried out in the university (nearly 70% of the production) as well as in the hospital (19%) sectors. In particular, with 2% of the output, the private sector produces very little in the field. Table 6 shows the index of specialisation by sectors and by province. The index of specialisation provides a mean to compare the production in a field, in this case genomics, to the global pattern of Canadian publications as indexed in the SCI. A score lower than one means less activity in genomics than is the case for the whole set of Canadian publications whereas a figure larger than one indicates a larger than usual output. To take an example, Table 6 shows that Canadian enterprises publish less in genomics than in all fields combined, which is a confirmation of the suspicion of a low level of scientific publications by the private sector. In fact, only Quebec's firms seem to be holding their own. This is probably a consequence of policies that aimed to develop biotechnologies in Quebec dating back from the mid-1980s as well as the large presence of pharmaceutical firms.

The relatively low scientific output of the private sector at the Canadian level indicates that there may not be so much advanced research being undertaken by the private sector in genomics. Potentially, this means that they may not be a fertile environment to absorb the university production in genomics and to turn these developments into commercial products. Hence it is suggested to investigate further the link between the university, the hospital and the private sectors and to see whether these links are sufficiently developed to foster the diffusion of scientific knowledge and findings in genomics towards the market. This could be done by way of questionnaires with the aim of investigating how university and hospital researchers diffuse their findings other than by way of publications and how this knowledge percolates into private firms.

Table 6 shows that only Quebec published more in genomics than in all fields combined. The table shows that the greatest contribution to Quebec's outstanding output is the hospital sector although her universities are also more active than is the case in other provinces. This is not surprising since a large part of genomics research is being undertaken in the health sector and that Quebec's health sector has a greater than average scientific output. This table also shows that smaller provinces have difficulties making a breach in the field of genomics. The evidence also suggests that the federal and provincial governments are less active in this field than in their global output. British Columbia is the only government that publishes significantly more in genomics than in all fields considered.

Table 6 Specialisation of Provinces in Genomics (1995-1997) by Economic Sector Relative to Canadian Production in SCI

Province	Economic sectors						TOTAL
	Entreprise	Federal	Hospital	Provincial	University	Other	
Alberta	0,35	0,30	1,10	0,17	1,07	0,31	0,94
British Columbia	0,39	0,40	0,85	1,55	0,85	0,73	0,83
Manitoba	0,65	0,37	0,91	1,07	0,98	1,14	0,86
New Brunswick	0,72	0,56	-	0,72	0,21	0,11	0,31
Newfoundland	-	0,27	0,45	-	0,67	0,57	0,59
Nova Scotia	0,19	0,40	0,55	-	0,78	0,37	0,68
Ontario	0,67	0,41	1,87	0,68	0,97	0,85	0,99
Prince Edward Island	-	0,20	-	-	0,70	-	0,50
Québec	1,00	0,86	2,16	0,36	1,24	0,43	1,32
Saskatchewan	0,41	0,95	0,29	0,51	0,62	0,56	0,66
TOTAL	0,66	0,48	1,76	0,78	1,00	0,67	1,00

Source : Observatoire des sciences et des technologies

It is interesting to compare the production of scientific articles to different indicators of input such as the general level of population by province, the percentage of active researchers as well as the percentage of grants. Table 7 shows that when one consider the production of scientific articles by researchers who have published 9 or more articles in genomics, Quebec leads the way but only slightly, with Ontario following closely.

In terms of output from the most active researchers relative to the share of the population of each province, Quebec once again outdistances other provinces while Alberta and Ontario hold their own share. British Columbia which performed well in terms of obtaining grants, do not produce as many articles as could be expected given its share of the Canadian population.

The smaller provinces do not produce all that much even when we factor in their smaller population. This might be a confirmation of the suspicion that a critical mass of research in genomics is not present outside Ontario, Quebec, Alberta and British Columbia. In fact, the amount of grants as well as the number of articles produced observed in the smaller provinces are so low that this precludes robust statistical inferences. As such, we will concentrate the remaining part of the analysis on the four leading provinces.

In terms of the productivity of researchers as measured by the percentage of articles divided by the percentage of researchers, Ontario and Quebec have an output that is proportional to their share. Other provinces score lower but these figures do not vary extremely from what could be expected given the provincial distribution of active researchers. In terms of percentage of production by the more active researchers as a function of the percentage of grants by province, Quebec leads the way by producing 1,3 times more than its share of grants. Ontario also produce slightly more than what could be expected if there was a direct correlation between grants and publications. Alberta holds its own as it seems to be doing across the indicators used here. British Columbia does not produce as much as we could expect given the amount of grants they had received during those years.

Table 7 Canadian Production in Genomics (1995-1997) Versus Various Measures of Input by Province

Province	Production by researchers with 9+ articles	%Production by researchers with 9+ articles	%Production/ %Population	%Production/ %Researchers	%Production / %Grants
Alberta	744	9,2%	1,0	0,9	1,0
British Columbia	609	7,5%	0,6	0,9	0,4
Manitoba	157	1,9%	0,5	1,0	0,6
New Brunswick					-
Newfoundland	34	0,4%	0,2	0,9	1,1
Nova Scotia	111	1,4%	0,4	1,1	0,4
Ontario	3162	38,9%	1,0	1,0	1,1
Prince Edward Island	10	0,1%	0,3	0,8	1,5
Québec	3217	39,6%	1,6	1,0	1,3
Saskatchewan	78	1,0%	0,3	0,9	0,5
TOTAL	8122	100%	1,0	1,0	1,0

Source : Observatoire des sciences et des technologies

2.4 PUBLICATIONS IN GENOMICS BY PLATFORM AND BY SECTOR OF USE

Nearly 42% of articles belong to the functional genomics platform. Other platforms such as genotyping (15%), new technologies (10%), sequencing (8%) and proteomics (6%) are not as important although still visible. An important number of articles cover multiple platforms (15%). These data suggest that generic technologies, such as the "new technology" as well as the "bioinformatics" platforms, may not be sufficiently developed in Canada considering the enabling role they play in providing the basic tools needed to develop further the field of genomics. However, it is hard to conclude that this is an exceptional case since we do not possess data to compare the output of Canada to that of other countries. These data would be useful to calculate the index of specialisation of Canada in different platforms and sectors compared to the entire scientific production indexed in the SCI for the field of genomics.

It is not surprising to find that nearly 65% of the production of scientific articles is oriented towards the health sector. Factors that contribute to this level of activity include the traditions in the field of genetics, one of the parent discipline of genomics, and the traditionally strong association between biology (which makes important contributions to genomics) and the health sector. Other sectors are not that important, each of them accounting for less than 10% of the output. Forestry and fisheries seem to be the poor relative of genomic research in Canada but the lack of comparable data with other countries means that one should be careful in the interpretation of these figures. These sectors seem to be underdeveloped considering their economic for Canada.

3.0 Conclusion

This bibliometric study shows that genomics in Canada is characterized by a high level of concentration. Most of the research is undertaken and it is mostly used in the health sector. The university and the hospital sectors contribute to a large part of the production while the private sector is not highly active in the field. In terms of technological platform, the field of genomics in Canada is highly developed in functional genomics but may not be sufficiently developed in enabling technologies (that is, bioinformatics and new technologies). Forestry, fisheries and agriculture do not account for an important share of research in genomics and this is somewhat surprising considering the economic importance of these sectors to the Canadian economy.

The leading provinces are by far Quebec and Ontario whilst Alberta and British Columbia follow at a distance. Quebec seems to provide a fertile environment that combines strong university research with higher than average involvement from the hospital and the private sectors. Many of these indicators converge to suggest that Quebec specializes more in genomics than other provinces although Ontario still leads in terms of absolute numbers in several areas. Alberta is consistently performing well suggesting it ranks third in genomics. British Columbia receives more grants but is not as productive as one could expect from this important input. Other provinces do not seem to have reached a critical mass whereby research in genomics could become self-sustaining and provide world-class results in important quantities.

Annex Keywords used to constitute datasets through keywords in title search

Where "*" is a wildcard representing any character

genom*, gene(s), sequenc*, receptor*, *DNA, *RNA, *genetic*, *genot*, *chromoso*', *cloning*, *transcript*, *expression*, *antisense*, *radiation hybrid map*, *physical map*, *linkage map*, *restriction map*, *candidate region*, characterization, *nucleoside*, *recombinant*, *microsatellite*, * yac *, * bac *, *meiotic*, *meiosis*, allele*, * est *, * snp *, *informed consent*, mutation*, *biostatistic*, *biocomputing *, *bioinformatic*, *biochip*, *haplotype*, *autosom*, exon *, intron *, *nucleotide*, * pna *, *nucleic acid*, *peptid*, *post-translation*, *phosphorilat*, *glycosylat*, *apoptosis*, * PCR *, *polymerase*, ploidy*, *2-hybrid*, *plasmid*, *minisatellite*, *transgen*, *loci *, *locus *.

French keywords used in addition to the above in the case of MRC and NSERC databases (which contain both English and French title)

génom, gène(s), *récepteur*, * séquenc*, *arn *, *adn *, *génétique*, *génot*, *caractérisa*, *clonage*, *nucléoside*, *méiotique*, *méiose*, * allèle*, allèle*, *consentement éclairé*, *biostatistiq*, *bioinformatique *, *nucléotide*, * apn *, *