available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Mapping the emergence and development of translational cancer research

Alberto Cambrosio[a,*], *Peter Keating*[b], *Simon Mercier*[b], *Grant Lewison*[c], *Andrei Mogoutov*[d]

[a]*Department of Social Studies of Medicine, McGill University (Montreal)*
[b]*Department of History, University of Quebec (Montreal)*
[c]*Evaluametrics (Richmond, UK)*
[d]*Aguidel (Paris)*

ARTICLE INFO

Article history:
Received 26 May 2006
Accepted 28 July 2006
Available online 31 October 2006

Keywords:
Neoplasms
Clinical cancer research
Translational research
Bibliometric analysis
Intercitation networks

ABSTRACT

Cancer research is one of the principal targets of translational research, yet the nature of the relationships between different forms of cancer research remains controversial. The paper examines publications in the cancer field during the 1980–2000 period. A network analysis software program was used to map evolving patterns of inter-citations between cancer publications, their different research levels and the transformation of their relational content. Both inter-citation and content maps provide striking evidence of the consolidation in the 1990s of a translational interface that was practically non existent a few decades before. In 1980, research was polarized according to the allegiance to either a clinical or a laboratory style. This same duality obtains in the year 2000, albeit with the additional presence of a third, *biomedical* player whose activities are similarly structured by a common orientation, rather than by an exclusive commitment to a specific sub-domain.

1. Introduction

The present paper introduces a new method of mapping cancer research and contributes new evidence to the ongoing discussion of two complementary issues:

- The first concerns the *policy* debate about the relationships between different forms of cancer research. Fueled by the perception that medical problems can be solved through biological innovation, public and private sponsors of research have designed *translational research* programs whose goal is to accelerate the exchange of concepts and techniques between fundamental biology and medical practice. Debates over the modalities of translational research have focused on whether it follows a unidirectional flow of information from the research laboratory to the patient, or whether there is in fact, as most authors claim, a two-way traffic between basic research and the clinic. According to Coleman and Harris, (p. 132)[1] for instance, the "bridge must be crossed in both directions, bringing concepts from the laboratory into the clinic and taking observations from the clinic to the laboratory". Description of the two-way traffic varies. Alternative representations describe translational research as a bridge connecting *two different worlds* "that only occasionally meet in an uneasy partnership" (p. 4211),[2] or as an emerging interface between the laboratory and the clinic (p. 43),[3] that should become *a distinctive sphere of*

* *Corresponding author*: Present address: Department of Social Studies of Medicine, McGill University, 3647 Peel St., Montreal (QC), Canada H3A 1X1. fax: +1 514 398 1498.
E-mail address: alberto.cambrosio@mcgill.ca (A. Cambrosio).
0959-8049/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2006.07.020

activity in its own right. Discussions of translational research frequently broach wider policy concerns including the now decade-long "crisis" in clinical research[4,5] and the ensuing need to reinforce this distinctive form of research by establishing dedicated institutions and career patterns.[6]

- Despite its evident policy implications, the second issue is primarily *historical* and *epistemological* and concerns the emergence in the late 20th-century of a distinctively *biomedical* form of cancer research. The term *biomedicine*, while now in common usage remains controversial. Ahrens (p. 34),[4] has dismissed it as an overly "inclusive word . . . for the many kind of research . . . whose content runs the gamut from strictly biological to strictly clinical". Perhaps more dramatically, when the president of the U.S. Biophysical Society argued in a 1999 *Nature* commentary that the central task of the biomedical enterprise was to "reduce the problems of disease to problems of molecular science", [7] he was sharply criticized for conflating "biology and medicine into an ill-defined hybrid 'biomedicine'".[8] One of the reasons for the persistence and resilience of debates about the nature of biomedicine is that while one can indeed speak of a realignment of biology and medicine based on the direct interaction of these two fields since WWII, their relations remain far from seamless and unproblematic.[9] Biomedicine handles the relations between the normal and the pathological not in terms of subordination (the reduction of the pathological to the normal) but in terms of mutual enrichment, leading to a complex interweaving of the diverse material and epistemic components of the life sciences.[10] Biomedical entities such as surface markers, oncogenes, and genetic signatures participate simultaneously in normal and pathological processes.

Our purpose here is not to propose further definitions of translational research, but to assess whether in recent years anything resembling a translational interface has emerged: we investigate the reality of a phenomenon, rather than attempting to shape its evolution. The emergence and development of biomedicine as a distinctive domain has been analyzed by several authors, using standard historical methods, such as archival work and oral history interviews.[9,11,12] While these methods offer a "thick description" of historical and social contingencies, their often-limited scope leave them open to criticism. We will thus revisit the issue by resorting to a semi-quantitative approach that involves the analysis of large publication data sets and the visualization of emerging patterns.

Previous contributions to the bibliometric analysis of cancer research were often limited to productivity measures.[13,14] New information visualization tools allow us to make complex relations and configurations visible without reducing them to a few statistical indicators, and thus also to analyze transformations in the organization and content of cancer research. In the present case, we will map the development of cancer research between 1980 and 2000 by examining both the evolving pattern of inter-citations between cancer publications and the transformation of the semantic networks that characterize these relations.

2. Materials and methods

2.1. Publication data

We obtained data concerning publications in the cancer field from the CD-ROM version of the *Science Citation Index* (SCI) produced by Thomson-ISI. The data included, in addition to bibliographic sources, the list of all cited references. From these data, we produced two databases:

- A first database contains the articles published in 121 journals specializing in cancer. One can safely assume that all these articles discuss cancer-related topics and that the database is thus highly specific. Sensitivity is, however, low: cancer journals publish only about 42% of all articles related to cancer.
- We thus created a second database that captures most cancer articles, regardless of the journal in which they were published, by exploiting a filter that uses title keywords (sometimes in combination) to select cancer articles in generalist publications.[15] The overall precision of the filter is 0.95 and its recall is 0.90.

The first database is asymmetrical, since it contains reciprocal citations between cancer journals but only unidirectional citations from cancer to generalist journals. The second database, in contrast, is symmetrical: *citing* and *cited* journals include both specialist and generalist publications.

2.2. Research levels

We assigned each journal in the two databases to one of the following four research level categories: clinical observation, clinical mix, clinical research and basic research. Research levels were derived from a combination of terms occurring in the titles of the articles they publish. They are dynamic, i.e., a given journal can switch category if its content is modified following, for instance, a change in editorial policy or the evolution of a journal's subfield. A detailed account of the research level algorithm is presented elsewhere.[16]

2.3. Mapping

We produced both inter-citation and semantic network maps using *ReseauLu* (http://www.aguidel.com), a network analysis software specifically designed for the treatment and mapping of heterogeneous relational data so that they can be visually inspected and interpreted.[17,18] *ReseauLu* uses a dynamic positioning algorithm simulating the interaction between objects. It does so through a three step optimisation process: (i) global initial positioning of the object vis-à-vis all the other objects in the space; (ii) micro-optimisation of the positioning of the object vis-à-vis the other objects to which it is directly connected ("network neighbours"); and (iii) meso-optimisation of groups of highly connected objects ("clusters"). The optimization process depends on explicit rules defining symmetry properties, structural equivalence of points inside the structure, centrality and "between-ness" of objects. The resulting map has no axes.

Given the high degree of inter-citation in the cancer field, we only mapped the more specific links. Specificity is here defined as the fact that a given journal preferentially cites articles in another journal, and is calculated by taking into account the number of potential citations (given the number of articles published by both the citing and cited journals) a given journal can, in principle, make to other journals. The size of nodes (journals) on inter-citation maps is proportional to the number of citations given and received by each journal: this indicator thus refers to the relative size (number of articles) and visibility (number of citations received) of each journal. Nodes are color-coded according to the research level of the journal.

2.4. *Text-mining*

To analyze the content (*semantic network*) of articles from selected key journals of different research levels we text-mined their titles and abstracts using *SPSS LexiQuest Mine* (http://www.spss.com/lexiquest/lexiquest_mine.htm). This software uses Natural Language Processing algorithms to extract single- and multi-word concepts from articles and calculates their co-occurrence within textual units (in this case, individual articles). The software uses dedicated dictionaries to extract only meaningful concepts, discarding uninformative terms.

2.5. *Semantic network maps*

These maps simultaneously display the links connecting the concepts (single and multi-word terms) that co-occur within the title and abstract of a given article, and the links connecting these co-occurring concepts to the journals from which they were extracted. The number of nodes that can be displayed on a map without affecting its readability amounts to approximately 200. To overcome, at least in part, this graphical constraint we used two distinct, yet complementary mapping approaches:

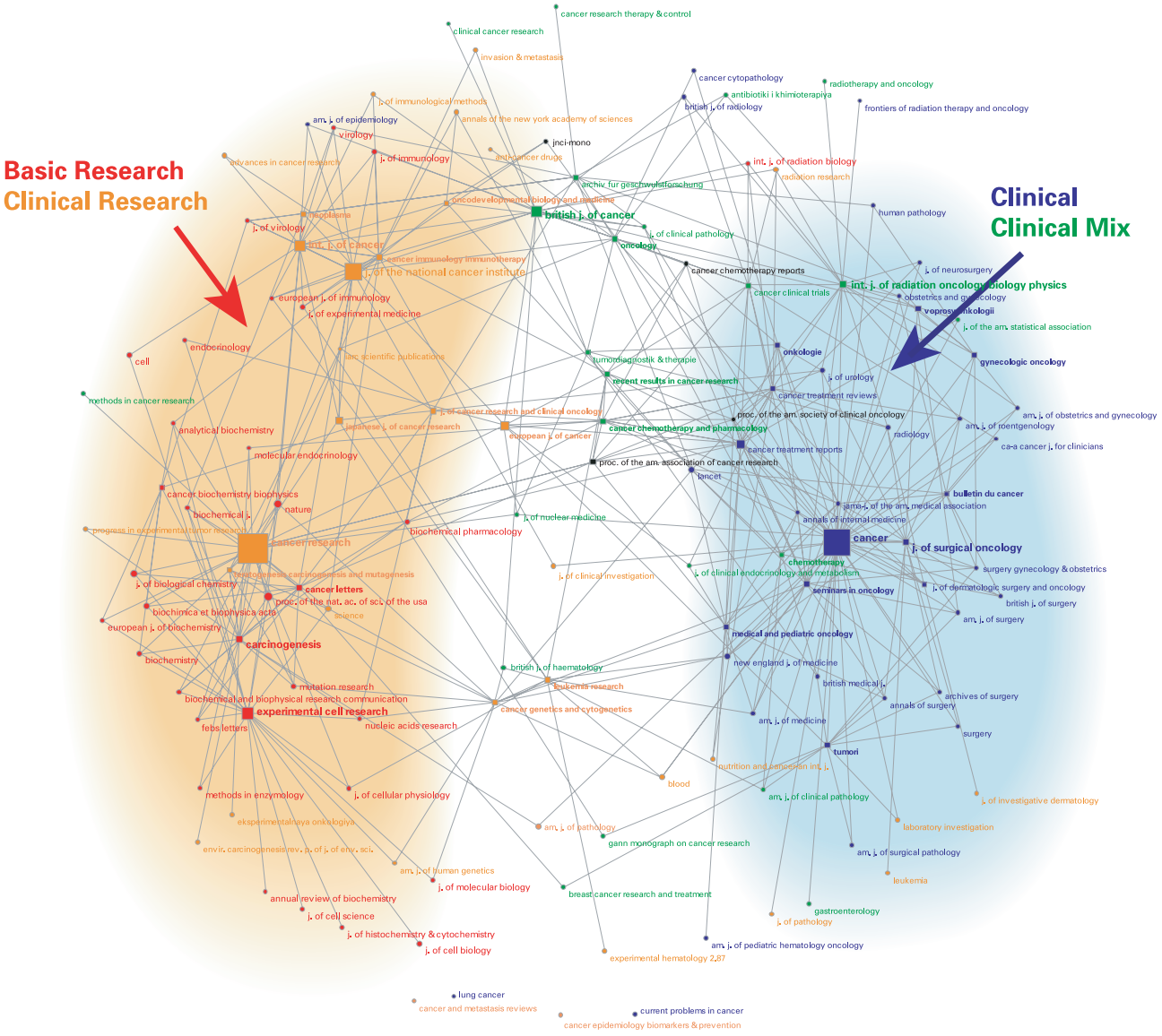

Fig. 1 – 1980 intercitation network of cancer journals. Specificity threshold 15% (see article for additional explanations). Journal color coding: red = basic research, orange = clinical research, green = clinical mix, blue = clinical observation, black: no research level available.

- A first approach consisted of selecting and mapping the most frequent (top 30) concepts found in each journal: this approach emphasizes the relations (or lack thereof) between journals as established by shared concepts;
- The second approach consisted of selecting the most frequent (top 200) co-occurring concepts in the entire database, and then situating the journals within the resulting, overall semantic network. To avoid size effects, we normalized the absolute frequency of terms extracted from different journals.[19]

3. Results

As explained in the *Methods* section, we analyzed two different databases, i.e., a first database with data concerning the citing behavior of the articles published in 121 cancer specialist journals, and a second database with data on cancer articles regardless of the journals in which they were published. The two databases clearly overlap. Given the already mentioned fact that graphical reasons dictate that the maximum number of journal nodes displayed on a map be limited to approximately 200, the resulting maps were not redundant:

- The first database provided information about inter-citation patterns from a specialist perspective. The 121 journals devoted entirely to the topic of cancer included both large, highly visible journals and small, less visible journals occupying specific niches in the field. All were displayed on the maps together with the names of the generalist journals they most often cite.
- The second database provided a more comprehensive view of the cancer field. So, for instance, while the 121 specialist journals included only one surgery journal (*Journal of Surgical Oncology*), maps derived from the second database display a cluster of generalist surgical journals. Similarly, although the first database already included multidisciplinary, basic science journals (*Nature*, *Science*, *Cell*, etc.) because the articles they publish are often cited by specialist journals, the maps derived from the second database

Fig. 2 – 2000 intercitation network of cancer journals. Specificity threshold 10% (see article for additional explanations). Journal color coding as in Fig. 1.

reinforced their presence to the point where they tended to form a discrete cluster. As a trade-off, small specialist journals were absent from these maps.

The maps derived from the two databases thus provided different, yet complementary perspectives on the cancer field. Our data covered the entire period from 1980-2003 but, given the large amount of data involved, we compared only selected years. In the present paper we report data only for 1980 and 2000, since intermediate maps display intermediate patterns.

Before comparing the maps, let us consider some of their general features. Figs. 1 and 2 display the inter-citation maps of specialist journals in the field of cancer for the years 1980 and 2000. They show only the more specific links using a 15% (1980) and a 10% (2000) specificity threshold. Fig. 3 displays the map of inter-citations of cancer publications in both specialist and generalist journals (10% specificity threshold). Each node corresponds to a journal title that appears only once on each map. The size of a journal's node is proportional to the number of citations given and received (or to the number of citations received in the case of generalist journals on Figs. 1 and 2), and the lines connecting the nodes correspond to citation links. Each node is color-coded: blue for clinical observation, green for clinical mix, orange for clinical research and red for basic research. A striking feature of the maps is that journals of a given research level preferentially cite journals of that same level. While certainly not a major discovery, this finding confirms the robustness of our approach, since research levels were calculated independently of the algorithms used to produce the maps.

Comparing Figs. 1 and 2 shows a first clear-cut trend: an increase in the absolute number of journals in the field and a diversification of the available publication outlets within each research level. Whereas in 1980 the clinical observation (blue) domain was dominated by a single journal, *Cancer* (established in 1948), in 2000 the *Journal of Clinical Oncology*, established in 1983 as part of the professional strategy of the American Society of Clinical Oncology,[20] now enjoys a similar visibility, with several other smaller

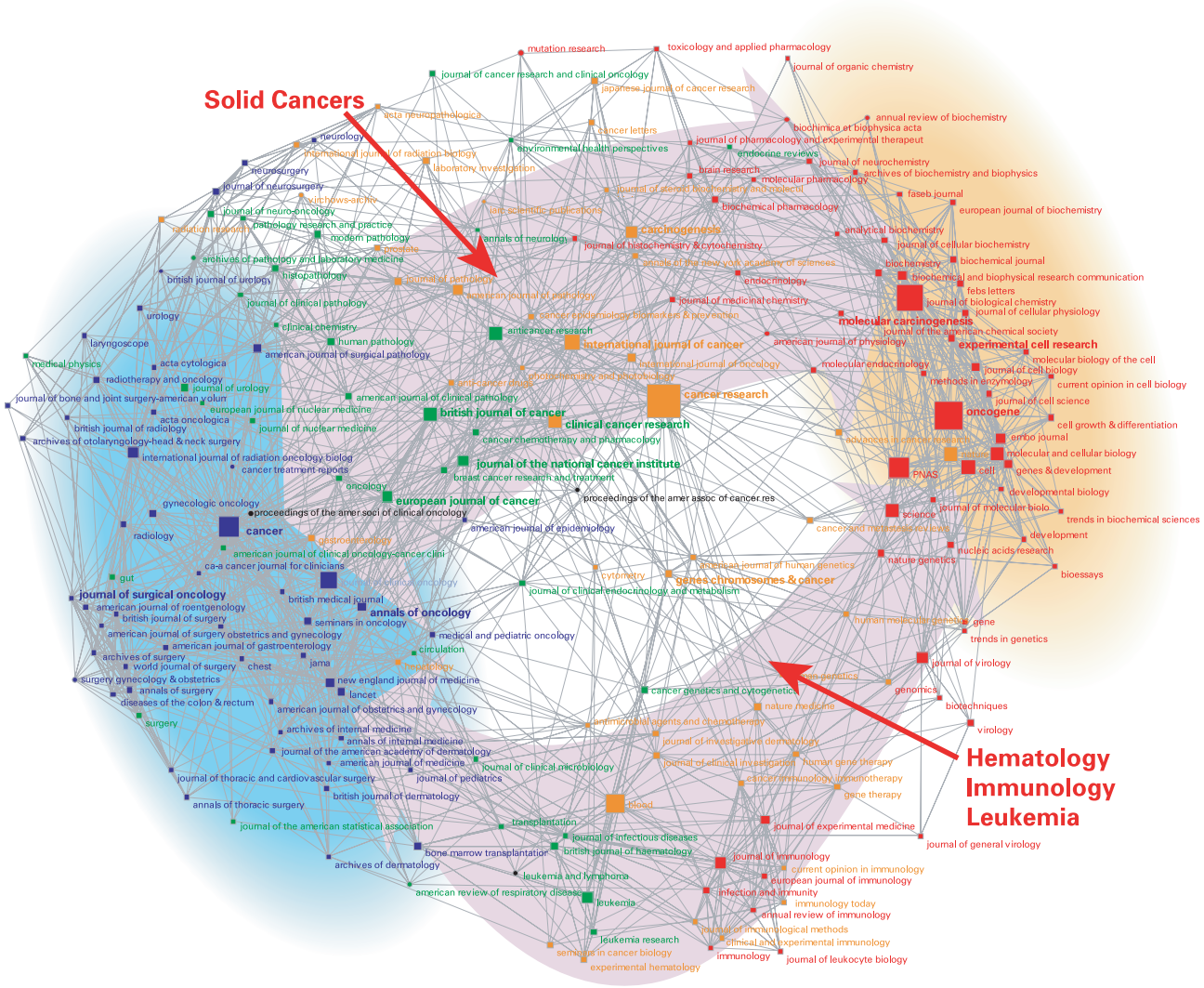

Fig. 3 – 2000 intercitation network of journals that published cancer articles. Specificity threshold 10% (see article for additional explanations). Journal color coding as in Fig. 1.

journals making their presence felt. In contrast, within clinical research (orange) publications, *Cancer Research* remains the strongest player from 1980 to 2000. As for the basic research (red) cluster, in 2000 it features the dominant presence of *Oncogene* (established in 1987), an obvious indication of the rapid rise and present fortune of molecular biological approaches in the cancer field: maps of previous years (not shown here) allow one to follow the emergence and establishment of a growing cluster of molecular biology publications: while such a finding is, once again, expected, the maps visualize this process over time. Similar findings concerning the role, position and changing fortunes of specific journals can be derived from a careful analysis of the maps, but the key finding emerging from a comparison of Figs. 1 and 2 is:

- In 1980 clinical journals (blue and green) and research journals (orange and red) sit at two opposite poles: there are only few inter-citational connections between these two poles, a result that tends to comfort the diagnosis of an occasional and uneasy partnership between clinicians and researchers.
- In the year 2000, we notice an interface between clinical observation (blue) and basic research (red) journals. Consisting of clinical mix and clinical research publications (green and orange), this interface evinces the emergence of a "translational research" domain that harbors specific *biomedical* activities predicated upon the alignment of normal and pathological investigations.

Fig. 3 displays the inter-citation map of specialist and generalist publications in the cancer field for the year 2000. As compared to Fig. 2, this map shows even more clearly a biomedical or translational interface. Obviously, most basic cancer research is published in generalist journals, such as the *Journal of Biological Chemistry*, *PNAS*, etc., and this results in a well-defined cluster of densely connected basic research journals. As previously mentioned, on the clinical observa-

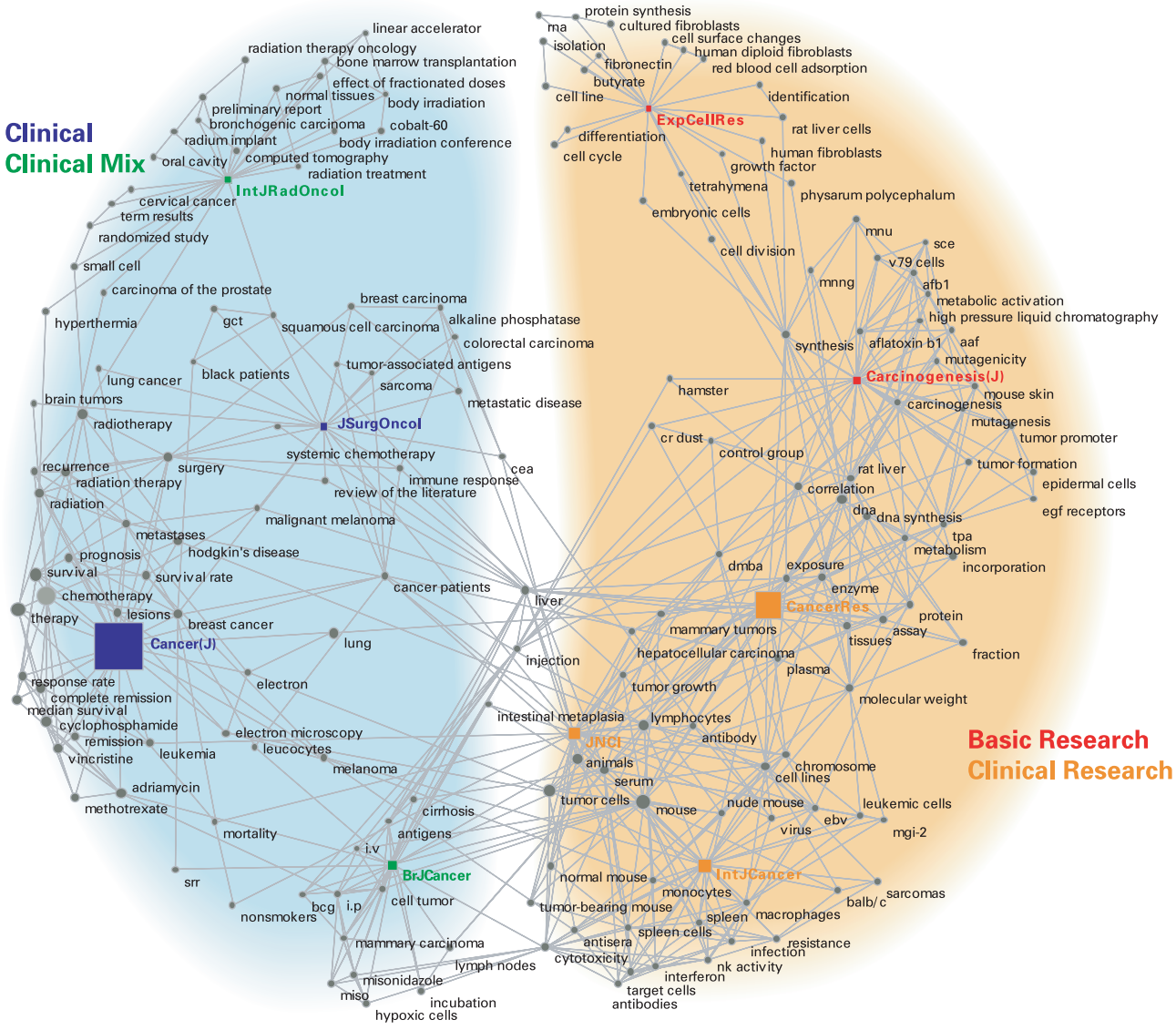

Fig. 4 – 1980 semantic network of selected cancer journals: top 30 concepts text-mined from each journal (see article for additional explanations). Journal color coding as in Fig. 1.

tion (blue) side we now find a cluster of surgical journals. As in Fig. 2, the clinical mix *cum* clinical research domain sits between these two poles, but Fig. 3 has the additional bonus of displaying two distinct translational paths connecting the clinical and basic research poles: at the top of the map, a path through publications on solid cancers, and at the bottom a path corresponding to onco-hematology.

So far, the maps have allowed us to investigate the transformation of inter-citation patterns. Do we find a similar trend if we look at semantic, rather than inter-citational networks? The answer is yes. Figs. 4–6 show semantic maps produced using the two different approaches described in the *Methods* section. In 1980 (Fig. 4) one finds almost no connections between co-occurring concepts extracted from journals at the clinical and the research poles, the exceptions being polysemic concepts such as *liver* or *injection*. The corresponding map for 2000 (Fig. 5) displays a semantic interface between the clinical and research pole linked to journals belonging to the translational research level (green and orange). This finding is confirmed by Fig. 6 that shows the semantic network for the year 2000 generated by the second mapping approach. At first sight, Figs. 5 and 6 look quite different, but closer analysis shows that the results are consistent: on Fig. 6, basic research (red) and clinical observation (blue) are situated at the opposite ends of the semantic spindle established by co-occurring concepts, while clinical mix (green) and clinical research (orange) journals have migrated towards the center of the web.

Compared to inter-citation networks, semantic maps offer even richer opportunities for analysis providing, as they do, insights into the conceptual and technical transformation of biomedical research activities. The analysis of the emergence of a translational interface connecting laboratory and clinical activities must accommodate the fact that terms such as "laboratory" and "clinical" do not have stable meanings: "clinical", which referred exclusively to the management of patients,

Fig. 5 – 2000 semantic network of selected cancer journals: top 30 concepts text-mined from each journal (see article for additional explanations). Journal color coding as in Fig. 1.

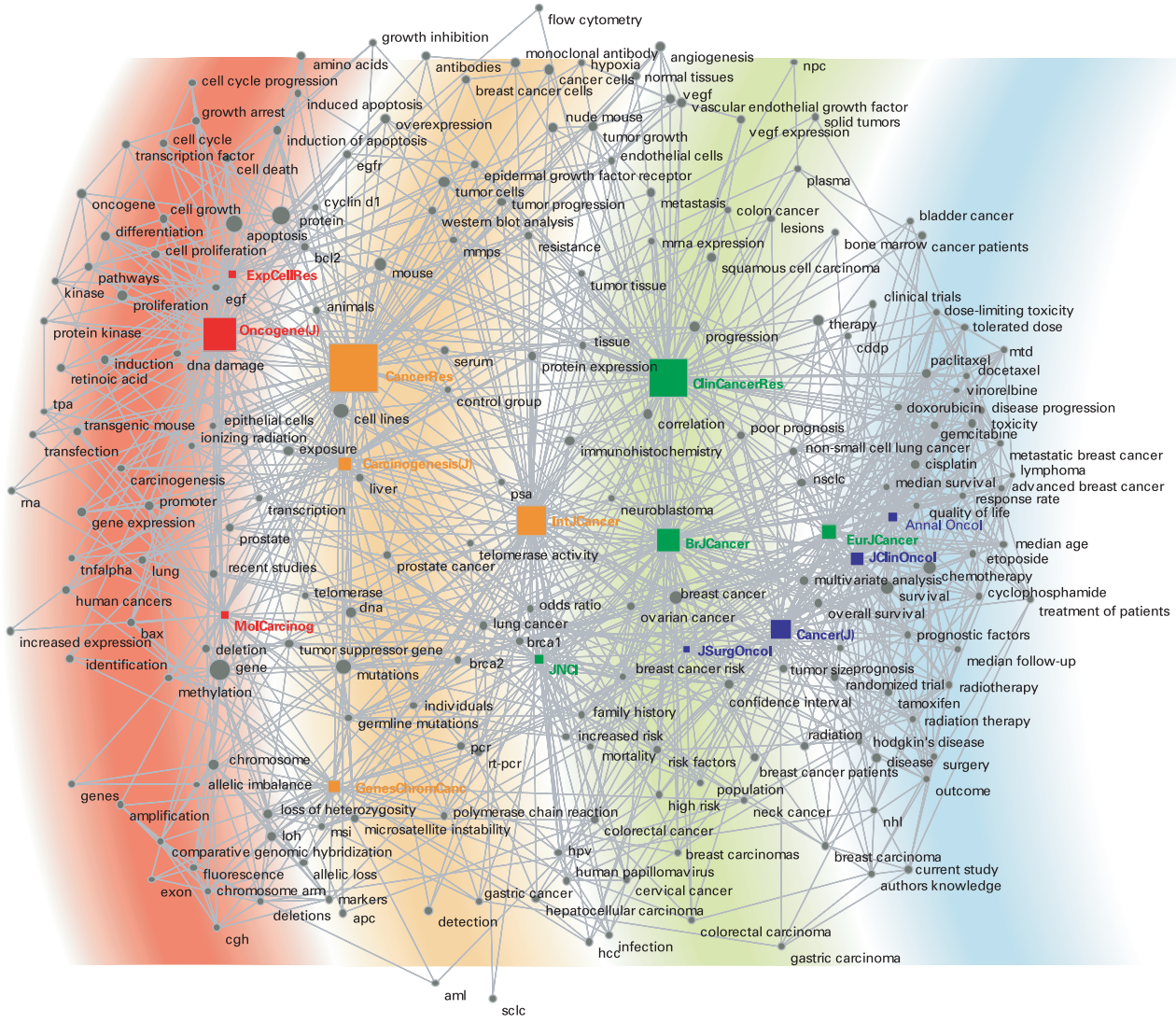

Fig. 6 – 2000 semantic network of selected cancer journals: top 200 concepts text-mined from the journal set (see article for additional explanations). Journal color coding as in Fig. 1.

now includes work on human material (such as human cells, genes, or antibodies). Conversely the notion of laboratory, no longer confined to animals, has come to include investigations on human samples. If we examine the concepts located in the middle section of Figs. 5 and 6, we notice terms such as *brca*, *mutations*, *genes*, *psa* that all refer to distinctively biomedical (i.e., simultaneously normal and pathological) entities, as well as terms referring to techniques and tools such as *monoclonal antibodies* and *rt-pcr* that are used to investigate both normal and pathological experimental systems. Given the limited scope of this article, we will not further analyze the content of our semantic maps, that were moreover designed to answer questions about the relations between research levels rather than to study the detailed conceptual structure of cancer research. We trust, however, that the discussion so far demonstrates the great potential of semi-quantitative mapping methods for the analysis of the interaction between the several components of a multifarious field such as cancer.

4. Discussion

Both inter-citation and semantic network maps of cancer research provide striking evidence of the consolidation in the 1990s of a biomedical interface that had barely begun to emerge a few decades before. Our maps show that in 1980, within the broadly defined field of cancer, research was less defined by adherence to sub-domains and more by a generic allegiance to either a clinical or a laboratory style. This same duality obtains in the year 2000, albeit with the additional presence of a third, *biomedical* player whose activities are similarly structured by a common orientation or style, rather than by an exclusive commitment to a specific sub-domain. Cancer policy analysts who frame their investigations in terms of pre-defined specialties or disciplines (such as biochemistry, genetics and the like) will likely miss the overall dynamics displayed by a network analysis of large domains.

Conflict of interest statement

No conflict of interest.

Acknowledgement

Research for this paper was supported by grants from the Canadian Institutes of Health Research (MOP-64372), the Fonds québécois de la recherche sur la société et la culture (ER-95786), the Social Sciences and Humanities Research Council of Canada (410-2005-1350), as well as by a CIHR/INSERM 2004-2005 International Exchange Award to AC. We would like to thank the "Observatoire des sciences et des technologies" in Montreal for giving us access to the SCI data.

REFERENCES

1. Coleman CN, Harris JR. Current scientific issues related to clinical radiation oncology. *Radiation Res* 1998;**150**:125–33.
2. Chabner BA, Boral AL, Multani P. Translational research: Walking the bridge between idea and cure. Seventeenth Bruce F. Cain Memorial Award Lecture. *Cancer Res* 1998;**58**:4211–6.
3. Pollak M. Peptide growth factors and breast cancer treatment: Recent translational research. *Recent Results Cancer Res* 1996;**140**:43–9.
4. Ahrens EH. The crisis in clinical research: Overcoming institutional obstacles. New York: Oxford University Press; 1992.
5. Schechter AN, Perlman RL, Rettig RA. Editors' introduction: Why is revitalizing clinical research so important, yet so difficult? *Perspect Biol Med* 2004;**47**:476–86.
6. Nathan DG. Careers in translational clinical research. Historical perspectives, future challenges. *JAMA* 2002;**287**: 2424–7.
7. Dill KA. Strengthening Biomedicine's Roots. *Nature* 1999;**400**:309–30.
8. Schechter AN. Medicine and biology are more than biomedicine. *Nature* 1999;**401**:424.
9. Keating P, Cambrosio A. *Biomedical platforms. Realigning the normal and the pathological in late-twentieth-century medicine*. Cambridge, MA: MIT Press; 2003.
10. Keating P, Cambrosio A. Does biomedicine entail the successful reduction of pathology to biology? *Perspect Biol Med* 2004;**47**:357–71.
11. Gaudillière JP. *Inventer la biomédecine: la France, l'Amérique et la production des savoirs du vivant (1945-1965)*. Paris: La Découverte; 2002.
12. de Chadarevian S, Kamminga H, editors. *Molecularizing biology and medicine. New practices and alliances, 1910s–1970s*. Amsterdam: Harwood Academic Publishers; 1998.
13. Ugolini D, Mela GS. Oncological research overview in the European Union. A 5-year survey. *Eur J Cancer* 2003;**39**:1888–94.
14. Grossi F, Belvedere O, Rosso R. Geography of clinical cancer research publications from 1995 to 1999. *Eur J Cancer* 2003;**39**:106–11.
15. Lewison G. The definition and calibration of biomedical subfields. *Scientometrics* 1999;**46**:529–37.
16. Lewison G, Paraje G. The classification of biomedical journals by research level. *Scientometrics* 2004;**60**:145–57.
17. Cambrosio A, Keating P, Mogoutov A. Mapping collaborative work and innovation in biomedicine: A computer-assisted analysis of antibody reagent workshops. *Soc Stud Sci* 2004;**34**:325–64.
18. Bourret P, Mogoutov A, Julian-Reynier C, Cambrosio A. A new clinical collective for French cancer genetics: A heterogeneous mapping analysis. *Sci Technol Human Values* 2006;**31**:431–64.
19. Widdows D. *Geometry and Meaning*. Chicago: The University of Chicago Press; 2003.
20. Krueger G. The Formation of the American Society of Clinical Oncology and the Development of a Medical Specialty, 1964–1973. *Perspect Biol Med* 2004;**47**:535–7.