Mapping the highly collaborative Stem Cell research field: Adding last-author-based analysis to the author co-citation analysis family

Dangzhi Zhao
School of Library and Information Studies
University of Alberta
dzhao@ualberta.ca

Andreas Strotmann
School of Public Health
University of Alberta
andreas.strotmann@ualberta.ca

ABSTRACT

In this paper we compare author cocitation analysis (ACA) results for the highly collaborative stem cell (SC) research field 2004-2009 using three types of ACA: all-author, first-author, and last-author. The latter of these, introduced here for the first time, is found to be an excellent compromise between first- and all-author ACAs in that (a) Scopus directly supports it and (b) its results are close to those of an (optimal) all-author ACA in fields where last authors are traditionally those who supervise the research published in a paper. We confirm predictions from previous studies that all-author ACA results have better model fits than single-author ACA ones, but cannot confirm the hypothesis that significantly higher levels of collaboration in a field lead to significantly greater differences between first- and all-author ACA results.

Keywords

Scholarly communication, author co-citation analysis, stem cell research.

INTRODUCTION

Despite caveats, citation analysis has been shown to be uniquely successful for studying the impact, structures and networks of scholarly communities due to its relative unobtrusiveness, objectivity, reliability, and low cost comparing to interview and survey techniques (Garfield, 1979; Harter & Kim, 1996; McCain, 1986; Small & Griffith, 1974; Smith, 1981; Sullivan et al., 1980; White, 1983; White, 1990; White & Griffith, 1981). Citation analysis methods and applications advanced dramatically through the citation indexes of the Institute for Scientific Information (ISI) which provided data and tools for large scale citation analysis studies, but they have also largely been limited by the incompleteness, bias, and other problems of these databases for citation analysis purposes (Edge, 1979; MacRoberts & MacRoberts, 1989).

Prominent among these problems is the fundamental limitation in author-based citation analysis theory and methodology that most studies have so far been limited to considering scholars' contributions as first authors because the ISI databases only index the first author of each cited reference. This has been consistently found problematic in at least some research fields when it comes to the evaluation of researchers based on citation counts (Lindsey, 1980; Long et al., 1980; Smith, 1981; Zhao, 2006a). On the one hand, first-author-based citation counting that these evaluations use disregards a researcher's contributions as non-first author. On the other hand, collaboration has become the norm rather than exception in many research fields (Cronin, 2007), which has been pushed even further by the fact that many if not all major funding agencies encourage collaborative research projects.

For evaluative citation analysis studies that attempt to map the intellectual structures of research fields through author co-citation analysis (ACA), however, findings regarding the effect of this problem have been inconsistent. Some studies found that all-author ACA produces clearer pictures of the research fields being studied that are easier to interpret in the computer science research fields (Zhao, 2006b; Schneider et al., 2007). Some studies showed that the intellectual structures revealed by the traditional first-author-based ACA and by all-author ACA are largely the same in the information science field (Persson, 2001; Zhao & Strotmann, 2008a).

These studies relied on highly specialized data sources on a small field (XML research) (Zhao, 2006b), or a single journal (IEEE transactions) (Schneider, et al., 2007), or incomplete data from the ISI databases or Scopus (Persson, 2001; Zhao and Strotmann, 2008a). Future studies were therefore suggested to use more reliable and complete data in order to test the hypothesis that this inconsistency in findings is due to the different collaboration levels in the computing science and information science fields and that the higher the collaboration level, the more pronounced the differences between first- and all-author ACAs may be

ASIST 2010, October 22–27, 2010, Pittsburgh, PA, USA.

(Zhao & Strotmann, 2008a).

The present study is such an attempt. It investigates the effect of different citation and co-citation counting methods on the results of ACA mappings of research fields through the study of a highly collaborative biomedical field, the stem cell research field. This is enabled by a large citation dataset we compiled from existing bibliographic databases through a sophisticated multi-step process that collects much cleaner data on stem cell research and much more accurate and complete information about cited references, compared to citation indexes commonly used in citation analysis studies, i.e. Scopus and the ISI databases.

Stem cell research has been rising to the forefront of biomedical science and public health and research policy in recent years.

"A stem cell is a special kind of cell that has a unique capacity to renew itself and to give rise to specialized cell types. …… Their proliferative capacity combined with the ability to become specialized makes stem cells unique" (Department of Health and Human Services, 2001). Stem cell research investigates the biological and medical promises of stem cells, with the long-term clinical goal of, on the one hand, improved understanding of cancers that develop from stem cells running amok, and on the other hand, utilizing the ability of stem cells to differentiate into a large variety of tissue types to assist in healing a wide range of "diseases, conditions, and disabilities including Parkinson's disease, amyotrophic lateral sclerosis, spinal cord injury, burns, heart disease, diabetes, and arthritis" (Department of health and Human Services, 2001).

Biomedical research fields, including stem cell research, are highly collaborative. According to Newman (2001), the mean number of authors per paper in the biomedical fields documented by MEDLINE was 3.75 and the average total number of collaborators per author was 18, compared to 2.22 and 3.59 respectively in the computer science fields. The collaboration level in some individual biomedical subfields was even higher (e.g., 6.18 authors/paper in the cardiovascular subfield) (Bordons, et al., 1996). Our data show that less than 10% of stem cell research publications in the past few years were single-authored, and about one in seven had more than eight authors.

Like chemists, biomedical researchers "work in individual laboratories in close association with their own group of students, postdoctoral fellows, and technicians" (Brown, 2010?, p. 307). The lab's head is often the principal investigator who develops initial ideas for research and procures the funding it requires. Its junior researchers often conduct the actual studies and perform the necessary experiments in the lab under the guidance of the lab head. Research results are published with that junior researcher as the first author, the lab head as the last, and the other lab members or outside collaborators involved in the research as authors in between (Sonnenwald, 2008, p. 670). This research and publication culture is clearly very different from what that commonly accepted in Library and Information Science (LIS) about the meaning conveyed by authorship order, and has not been taken into account sufficiently in citation analysis studies due partly to the limited support provided by available citation indexes.

The present study therefore introduces last-author ACA and explores what it means and how it compares with first- and all-author ACAs in a field where last authors traditionally play a special role in the genesis of a paper.

Specifically, the present study maps the stem cell research field, taking into account its highly collaborative nature and unique research and publication culture. It focuses on the following research questions:

1. How is ACA mapping of research fields affected by citation and co-citation counting methods?
2. How does last-author-based ACA compare with first- and all-author ACAs?
3. Do differences between single- (e.g., first-) and all-author ACAs change with the collaboration level of the research field?

Results from this study should contribute to a more thorough view of the theory and methods of citation analysis which has been central to the quantitative study of science and to LIS. The methodology developed and used in this study for collecting and analyzing citation data on all contributions of scholars can be applied directly to the study of a much broader range of research fields, namely all of bio-medical research, and can readily be adapted to any other highly collaborative research communities.

METHODOLOGY

In order to address these research questions, we will first examine results from all-author counting using fractional counts – the most preferred counting method for allocating credit in the case of multi-authored works (Lindsey, 1980; van Hooydonk, 1997; Zhao, 2005; 2006a). We will then compare these results with those from complete all-author counting, from last-author counting and from the traditional first-author counting.

We will use exclusive all-author co-citation counting to calculate co-citation matrices whose diagonal values are the authors' exclusive co-citation counts with themselves, a method that has been shown to be the most preferred both theoretically and in practice (Ahlgren, Jarneving & Rousseau, 2003; White, 2003; Zhao & Strotmann, 2008a).

To clarify, when an article by N authors is cited, each of these N authors' fractional citation counts increase by 1/N ; their complete citation counts increase by 1; and only the first or the last author's citation count increases by 1 when first- or last-author counting is used. The exclusive co-

citation count of author A and B increases by 1 whenever a paper cites at least one paper from A's oeuvre and at least one *different* paper from B's oeuvre. An author's oeuvre is defined as the collection of all papers written by this author as first author in case of first-author counting, as last author in the case of last-author counting, or as an author listed in any position in the byline in the case of all-author counting, resp. Clearly then, the only difference between fractional and complete all-author ACA is how authors to be included in ACA are selected based on their citation rankings using either fractional or complete citation counting.

Data Collection

In order to study the scholarly communication patterns of a research field using a citation-based approach, a set of publications in this field during a certain time period needs to be collected to represent this research field. The scholarly communication patterns of this field can then be studied based on the perceptions of authors of these publications as expressed in their citation behaviors recorded in citation links they provide in these publications. Clearly, the more complete and clean this set of publications is (i.e., including as many papers as possible on this research field and as few as possible on research outside of this field), the better a research field is represented and therefore the better its scholarly communication patterns can be studied. The citation links in these publications are an essential part of the dataset, and a complete list of authors of each cited reference should be included in order to take into account all contributions of the authors regardless of their positions in the by lines.

During the last few years, between around ten thousand publications per year have been published in the stem cell research field. Less than 10% of these publications were single-authored, and one in seven had more than eight authors. Given the magnitude of the dataset, the pervasiveness of multi-authorship, the multidisciplinary nature of this field, and limitations of current citation databases (i.e., Scopus, ISI citation databases), traditional core journal- or keyword search based methods using existing citation databases do not work well for this study for a number of reasons. We therefore developed and employed a multi-step process to build a dataset that is close to complete, clean and accurate compared with a dataset directly from existing citation databases. Details of these reasons and the steps and algorithms of this process can be found in Strotmann, Zhao, & Bubela (2010). A summary of the key points is provided here.

Limitations of Existing Citation Databases for Studying Highly Collaborative, Multi-disciplinary Research Fields

(a) We suspect that the highly collaborative nature of the stem cell research requires all-author counting for ACA purposes, which requires a complete list of authors of each cited reference. ISI citation databases only index the first author of a cited reference and Scopus provides up to eight authors. Scopus may be good enough for research fields such as Library and Information Science, but does not suffice for the highly collaborative stem cell research field in which there are many papers with more than eight authors.

(b) The stem cell field is highly multidisciplinary, with research ranging from biology to therapy, across all organs to a variety of diseases, and from biomedical sciences to social sciences and law. Journals that publish stem cell research are highly diverse on the one hand, and cover non-stem cell research extensively as well on the other.

(c) The stem cell field is large and extremely fast-growing. The number of publications within a year in this field is already beyond the limit that Scopus put on search results for download (i.e., 2000). Refining the search by journal does not work for reasons in (b).

Creation of a Complete and Clean Dataset of Stem Cell Research

1) We used a MeSH term search on "stem cell" in PubMed, and limited the search by document type to only include research articles and reviews. We selected a citation window of six years from 2004 and 2009 .

The actual searches for the years 2004-2007, 2008, and 2009 were carried out in December 2008, August 2009, and May 2010 respectively to allow sufficient time for PubMed to index the papers. A total of 31 040 papers was retrieved.

2) We created a set of search strings from these PubMed records, and issued these search strings in Scopus manually in order to retrieve these papers along with their cited references. About 98% of the papers were found in Scopus, and were subsequently kept in the dataset for our study.

3) 2 281 584 (or 95%) of the 2,405,522 cited references were found in PubMed. Those that were not found there were added by parsing the cited reference information, which includes the names of up to eight authors.

Data Analysis

Author Name Disambiguation

In a highly diverse and multidisciplinary field like stem cell research, the problems with author names (e.g., spelling variations of the same names, same author with different names and same names for multiple authors) are extremely pronounced. Author name disambiguation therefore became a necessary component of author-based citation and co-citation counting.

We summarize below the key points of the method we used for author name disambiguation. Details of this method can be found in Strotmann, Zhao, & Bubela (2009).

Citation and Co-citation Counting

We ranked cited authors by the number of times they are cited by papers in our dataset based on first, last and all-author counting (both complete and fractional). The top 300 authors of each of the four original rankings were selected for an author co-citation analyses (ACA), and their co-citation counts were calculated and put into a matrix. The diagonal values of the co-citation matrices are authors' co-citation counts with themselves based on the corresponding counting methods. For example, in the last-author co-citation matrix, the diagonal value of author A is the number of papers that cited at least two different papers written by author A as the last author. Clearly, the diagonal value of an author in the all-author co-citation matrix is normally larger than the sum of this author's diagonal values in the first and last-author matrices.

Factor Analysis and Visualization

Each of the four author co-citation matrices was factor analyzed and the results visualized as described in Zhao & Strotmann (2008b). We only included the top 200 authors in each of the 300x300 matrices in the factor analysis to meet the variable-case ratio requirement by the factor analysis method (Hair, 1998). Considering the size of the stem cell field, we chose a threshold here (i.e., 200) for author selection that is much larger than the largest seen in published ACA studies to date (White & McCain, 1998).

To compare results from different citation and co-citation counting methods, several relevant features of the author maps that they reveal were examined, including which specialties are identified, which specialties are most active, how these specialties are related to each other, how clear the specialty structure is, how well the research of individual scholars is recognized in the structure, and which are the central, peripheral or bridging specialties and scholars (White & McCain, 1998; White, 1990).

RESULTS

Factor Model Fit

Table 1 shows the factor models produced by the factor analysis routine in SPSS 7.0 based on Kaiser's rule of eigenvalue greater than 1, along with their model fits. For example, a factor analysis of the first-author cocitation matrix resulted in a 20-factor model which explains 88.3% of the total variance, and the differences between observed and implied correlations were smaller than 0.05 for the most part (almost 100%).

Table 1 confirms findings from previous studies (Zhao, 2006b; Zhao & Strotmann, 2008b) that first-author-based ACA represents a more fragmented picture of the field compared to all-author counting, especially complete counting, due partly to the fact that all-author counting favors highly cited research groups, resulting in top cited authors included in the ACA being less diverse. Last-author ACA produces an even more fragmented picture.

The higher concentration of the picture produced by all-author counting can be seen from the many highly successful research groups in the specialties identified through all-author ACA. For example, Nadal-Ginard, Anversa, Leri, and Kajstura published together frequently on myocardial regeneration, and all top 5 highly cited papers by each of these authors are co-authored by all four authors. These papers were published in top journals for biomedical research such as Nature, Cell and New England Journal of Medicine, and have been cited 874, 476, 378, 202, and 163 times respectively. Another example is Jeffrey Isner's team who published highly cited papers in journals such as Science and Circulation Research. Asahara and Silver appeared to be long time core members of the team and were coauthors on all top 5 publications. They were joined by a few other researchers at different times for different publications (e.g., Murohara, Li, Masuda, Kearney, Kalka, Takahashi).

Input co-citation matrix	# factors	Total variance explained	#(%) |non-redundant residuals| > 0.05
First-author	20	88.277%	123 (0%)
All-author (fractional)	19	92.626%	96 (0%)
All-author (complete)	19	95.334%	38 (0%)
Last-author	23	89.582%	132 (0%)

Table 1. Factor models and their model fits

Complete all-author counting brings this tendency to the extreme in that authors of a single highly cited paper were separated out into their own factor. For example, the seminal paper by James A. Thomson and colleagues at the University of Wisconsin-Madison (Itskovitz-Eldor, Shapiro, Waknitz, Swiergiel, Marshall, Jones) on "Embryonic Stem Cell Lines Derived from Human Blastocysts" published in Science in 1998 has been cited 1879 times, and its authors form a separate factor that is highly connected to the rest of the research area on human and other embryonic stem cells. Similarly, the paper by Pittenger and colleagues at Osiris Therapeutics (Mackay, Beck, Jaiswal, Douglas, Mosca, Moorman, Simonetti, Craig, Marshak) on "Multilineage Potential of Adult Human Mesenchymal Stem Cells" published in Science in 1999 started the area of mesenchymal stem cells and regenerative medicine. It has been highly cited (2511 times), and its authors were separated out into their own factor from all other studies on mesenchymal stem cells and regenerative medicine.

Overall Structure of the Stem Cell Research Field

Figures 1-3 are visual representations of the factor analysis

results from fractional all-author counting, first-author counting, and last-author counting. Results from complete all-author counting are not shown here because they agree quite well with fractional all-author results.

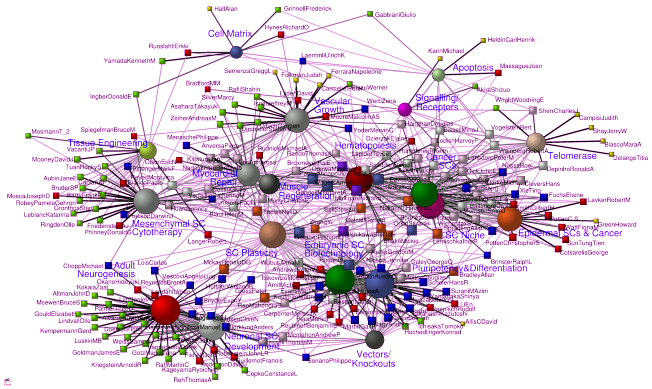

Figure 1. Researchers, specialties and their interrelationships – All-author ACA, fractional ranking

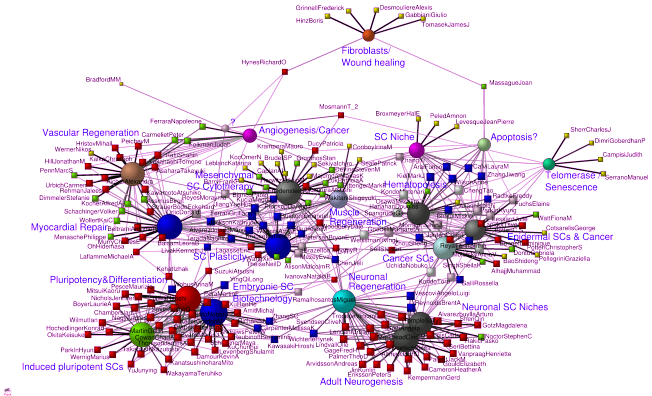

Figure 2. Researchers, specialties and their interrelationships - First-author ACA and ranking

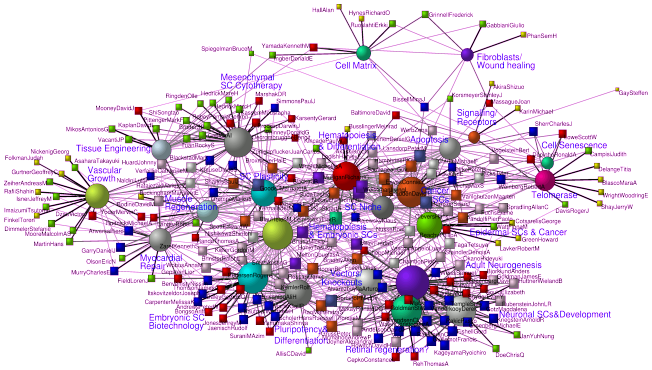

Figure 3. Researchers, specialties and their interrelationships – Last-author ACA and ranking

In these figures, the circular nodes represent specialties and the square nodes authors. The size of a specialty node is accumulated from loadings and serves as an approximate indicator of its overall significance in the map. The color of an author node indicates the number of specialties in which this author has membership: yellow for authors who only have membership in a single specialty, green for two, red for three, blue for four, and other colors for more than four specialties. The width and the greyscale value of lines connecting nodes are proportional to the value of the author's loading on this factor and represent the degree of relatedness, with darker and thicker lines representing closer ties. Only loadings that are 0.3 or higher are counted as in White & McCain (1998).

Table 2 provides the label and size of each of the factors in the three sets of results. The size of a factor is the number of authors who primarily load on this factor. The label of a factor is given based on an examination of the highly cited papers written by authors who load highly on this factor. The highest loading in each factor is also listed in the table as an indicator of clarity and distinctiveness of a factor.

All three maps show horizontally across their centers a loosely connected arc of specialties in the stem cell research field that, broadly, appear to focus on medical implications and applications of stem cell research. One half of that arc can be categorized as regenerative medicine, which aims to utilize the potential of stem cells to grow new tissue for repair or other treatment. The other half of the arc focuses on cancer research, where the goal is to understand and control the proliferative potential of cancer stem cells. The two halves are generally bridged by research on haematopoietic (i.e., bone marrow) stem cells, which has strong connections to (blood) cancer on the one hand but also provides a central ingredient for transplantation-style regenerative medicine.

Research on the (re-)growth of blood vessels has an obvious connection to regenerative medicine, where vascularization of new tissue is a universal requirement, and a less obvious one to cancer medicine where inhibition of neovascularization of cancerous tissue is a possible target for treatment. This area serves as a second bridge between the two parts of the central arc in the case of fractional ACA, whereas it splits into two separate but well-linked specialties (one regenerational, one cancer-related) whose links provide the second bridge. In the case of last-author ACA, it is visualized at the far regenerative end of the arc.In two ACA maps, a large central node connects this central arc to two clusters of highly interconnected research specialties below, one on neurogenesis and one on embryonic and pluripotent stem cells. The connecting central node itself appears to represent the idea that stem cell research has a common theme, namely, the *plasticity* of stem cells, i.e., their ability to differentiate into any type of body tissue. On the last-author ACA map, this factor is smaller and embedded at the center of the central arc.

Across the top of the figure, and more or less loosely connected to the different parts of the central arc, we see a

very loosely connected outer arc of small specialties. On the cancer end of the map, this includes research on cell senescence and cell death (which stem cells, like cancer cells, are able to avoid) and on extracellular regulation of stem cell differentiation and maturation.

	First		All		Last	
Factor	**Size**	**High load**	**Size**	**High load**	**Size**	**High load**
Adult Neurogenesis	17	1.05	31	1.06	25	1.03
Neuronal Regeneration	9	0.73				
Neuronal SC Niches	5	0.89				
Mesenchymal SC Cytotherapy	24	0.99	19	0.99	20	0.96
Embryonic SC Biotech.	20	0.97	19	0.95	18	0.98
Pluripotency & Differentiation	9	1.02	18	0.90	9	0.81
Induced Pluripotent SCs	12	0.86				
Vascular Growth			17	0.99	16	1.01
Vascular Regeneration	16	1.03				
Angiogenesis / Cancer	2	0.82				
Cancer SCs	17	0.99	14	0.89	17	0.88
Hematopoiesis	14	0.96	14	0.81		
Hematopoiesis& Embryonic SCs					4	0.68
Hematopoiesis & Differentiation					10	1.06
Neuronal SC Development			10	1.01	7	0.86
Epidermal SCs & Cancer	6	0.99	8	1.03	6	1.04
Telomerase			8	1.02	8	1.05
Telomerase / Senescence	4	0.97				
Cell Senescence					1	0.5
Myocardial Repair	12	1.01	8	0.84	5	0.86
Muscle Regeneration	3	0.99	7	0.87	5	1.03
Fibroblasts / Wound Healing	6	0.99			4	0.96
Retinal Regeneration					6	0.99
Cell Matrix			6	0.90	5	0.88
Signaling / Receptors			4	0.85	4	0.91
SC Niche	3	0.98	4	0.81	3	0.52
Tissue Engineering			4	0.79	5	0.96
Apoptosis	2	0.48	3	0.93	4	0.60
UD - Stem Cell Plasticity	16	1	3	0.58	9	0.76
UD - Vectors / Knockouts UD	2	0.51	2	0.57	7	0.78

Figure 2. Factor labels, sizes, and highest loadings

On the regenerative end, we see wound healing, tissue engineering and cell/matrixinteraction as peripheral specialties. Different ACA maps all show this outer arc, but with different subsets and combinations of these specialties.

Major Clusters of Specialties

Embryonic / Induced Pluripotent Stem Cell Cluster

This cluster of research specialties deals primarily with totipotent embryonic and pluripotent adult stem cells (which can grow into any kind of tissue).

In results from fractional all-author counting, it consists of two large and highly correlated research specialties (Embryonic SC Biotechnology, Pluripotency & Differentiation) and a small specialty (Vectors/Knockouts) that connects this cluster with the Neuro-cluster described below. The small specialty is missed only by first-author ACA, which identifies a third large specialty (Induced Pluripotent SCs) not recognized by the other ACAs in this cluster. Last-author ACA shows all three specialties identified in fractional ACA, and adds Haematopoiesis & Embryonic SCs as a specialty bridging to the central medical applications arc. Highly influential embryonic SC researchers are Thomson and Itskovitz-Eldor who together with their colleagues first successfully isolated embryonic stem cells from human embryos and grew them in culture. Their seminal paper was published in Science in 1998, and has been cited 1879 times. They are ranked 9 and 18, resp., among all stem cell researchers by fractional counting.

Most influential pluripotent SC researchers are Jaenisch, Yamanaka, and Takahashi, ranked 8, 12, and 38, resp., among all stem cell researchers. Yamanaka and Takahashi co-authored several highly cited papers, the most famous of which (cited 839 times) is their 2006 paper in the journal Cell, reporting their discovery of how to induce pluripotent stem cells from mouse embryonic and adult fibroblast cultures. This discovery started a whole new fast-growing research area in the stem cell field. The highly influential researchers identified in the other ACAs are largely the same as in fractional ACA. Only in the first-author ACA results do we see a major difference in that they only identify the first authors of the extremely influential papers (e.g., Thomson and Takahashi rather than Itskovitz-eldor and Yamanaka). Researchers who do not have an extremely highly cited paper but are nevertheless highly influential overall are not identified by first-author counting (e.g., Itskovitz-Eldor and Jaenisch).

Neuronal stem cell cluster

This cluster generally consists of two large and very strongly interconnected research specialties, one on Neuronal SC Development and one on Adult Neurogenesis. Last-author ACA factors out a small research area focused on Retinal regeneration from the latter field, whereas first-author ACA appears fails to identify the former and instead factors out research on Neuronal SC niches on the one

hand, and on Neuronal regeneration on the other.

The most influential researcher in the Neuronal SC development area appears to be David Anderson at Howard Hughes Medical Institute, ranked 35. In the Adult neurogenesis field, the most influential researchers are Gage, Alvarez-buylla, Weiss, Reynolds, Rakic, and Garcia-verdugo, ranked 2, 4, 14, 16, 22 and 24 respectively. These authors are identified in all ACAs.

Regenerative Medicine

This part of the central medical applications arc of specialties is divided largely by the class of stem cells involved – mesenchymal stem cells in one specialty (Mesenchymal SC cytotherapy), myocardial stem cells in another (Myocardial repair), muscle stem cells in a third (Muscle regeneration), and endothelial progenitor cells in a fourth (Vascular growth). The first of these targets cytotherapy of the inner organs; the second focuses on repairing infarcts; the third, on repairing muscle tissue; and the fourth, on regrowing or repairing blood vessels. Specialties in this cluster are much more loosely connected compared to the embryonic SC and the neuro-SC clusters. Only in the first-author ACA map does the Vascular growth specialty, which bridges this part of the central arc and the other, cancer research, part of the arc, separate out into a Vascular regeneration and an Angiogenesis/Cancer factor.

The Mesenchymal stem cell cytotherapy specialty is the largest in the cluster. Prockop, Caplan, Verfaillie, and Pittenger are the most influential researchers here, ranked 3, 7, 26, and 30, resp., among all stem cell researchers. The abovementioned seminal 1999 Science paper by Pittenger and colleagues at Osiris Therapeutics marked the start of this area of research, and has been extremely highly cited (2511 times).

The 2nd largest in this cluster is the Vascular growth / regeneration / Angiogenesis one. Dimmeler, Asahara, Rafii, and Isner are the most influential here, ranked 25, 37, 39 and 51, resp. Isner, Asahara and colleagues co-authored several highly cited papers published in journals such as Science and Circulation Research. Their 1997 Science paper on "Isolation of putative progenitor endothelial cells for angiogenesis" has been cited 1110 times.

The highly influential researchers identified are the same in all three sets of results except that first-author counting did not identify Verfaillie as well as Isner who is the lab head but is less highly ranked by fractional all-author counting.

Cancer Research

While it has long been understood that stem cells and cancer cells have much in common (in particular, their immortality and their ability to proliferate), the discovery that cancer stem cells (i.e., stem cells running amok) are primary initiators of tumours has greatly added to the interest in stem cell research among cancer researchers.

Research on Cancer SCs is therefore a central specialty of the cancer part of the central medical research arc in all ACA maps. This subarea is particularly active with respect to leukemia (cancers of the blood), which connects it directly to the haematopoietic stem cell research area.

The second major type of cancers being studied extensively is skin cancer, and Epidermal (or epithelial) SCs are instrumental to this specialty found in all ACA maps.

A third specialty found on all maps of this part of the stem cell research field studies the concept of a stem cell niche – the idea that the cells surrounding a stem cell determine the type of cells that it will differentiate into. Apart from differences in sizes, this separation is common to all maps.

The most influential cancer researchers in the stem cell field are Morrison, Dick, Clarke, Fuchs, and Watt, ranked 6, 11, 15, 20, and 28, resp. The former three are in the Cancer SCs specialty, while the latter two represent the Epidermal SCs & cancer specialty. They are the same in results from all ACAs, except that the first-author ACA did not identify Dick and Clarke.

Haematopoietic Stem Cells

The study of bone marrow (haematopoietic) stem cells forms a large specialty that connects regenerative medicine with cancer research (in particular, leukemia) as well as with research on embryonic stem cells in all maps.

In fractional ACA, this specialty is not distinct with mostly medium to low loadings. The most influential researcher in this area appears to be Orkin (ranked 17) who is also identified by last-author ACA, but not by first-author ACA.

The last-author ACA results see this area constituted of two sub-specialties, one of which focuses on early embryonic development of these cells, and the second of which concentrates on the later differentiation of cells from them. The first-author ACA results, on the other hand, show this research field as part of cancer research rather than bridging to regenerative medicine.

Integration of Stem Cell Research

At the center of Figure 1 we see a factor with a large number of low-to-medium co-loadings with (almost) every other factor on the map, and without any high author loadings that would give it a defining core. The papers that constitute this factor can only be characterized as spanning the breadth of the entire field – all the organs (liver, muscle, brain, bone marrow) which characterize individual clusters of the regenerative medicine and cancer related research, and both biology and medical applications of stem cells.

We suspect that this factor represents studies that attempt to construct a cohesive landscape of stem cell research from across its separate research areas, and thus to integrate findings from across the main dimensions of the field (embryonic and pluripotent vs more specialized stem cells;

medical applications that induce vs inhibit stem cell proliferation and differentiation) into a coherent picture.

This interpretation is confirmed by those authors (like Blau) that do show high loadings on a factor with the same label at the center of Figure 2 (first-author ACA). Figure 3 (last-author ACA) also shows this factor, but not in such a central position.

Peripheral Specialties

The intellectual landscape of the stem cell field is rounded off by a number of smaller specialties, which form a wide arc across the entire upper part of all maps, loosely linked to the central medical research arc. Different ACAs identify different combinations of specialties here, but common themes are quite apparent.

The factors that connect to the cancer research side of the central arc deal with different phases of the stem cell life cycle – cell maturation (and its coordination with the surrounding Cell matrix via Signaling/Receptors), cell senescence (or the lack thereof in stem cells due to Telomerases), and cell death (Apoptosis). Only one factor representing two of these themes is distinct in the first-author ACA, and the last-author ACA separates out an indistinct additional factor from the Telomerase one.

The peripheral specialties that are linked to the regenerative medicine aspect of stem cell research, are concerned with Fibroblasts/Wound healing, with the biology of the attachment and adhesion of (stem) cells to the surrounding Cell matrix, and with the biotechnology involved in mimicking the three-dimensional cell matrix and both its biomolecular and physical-mechanical properties for ex-vivo Tissue Engineering. These specialties are connected to each other, and to the peripheral cancer-related specialties.

The last-author ACA identifies all three of these themes. First-author ACA shows only one factor, Fibroblasts / Wound healing, which is missed by the fractional ACA.

Massague (ranked 23) loads high on Signaling/Receptors in this part of the maps.

DISCUSSION

Interestingly, last-author ACA results are very similar to fractional all-author ACA ones, much more similar than first-author ACA results are. This can be seen from both the structure of the field (overall, major clusters of specialties, and peripheral specialties) and the most influential researchers identified.

The major difference, which first-author ACA shares, appears to be that last-author ACA sometimes produces two smaller factors which stay in a single specialty in the all-author ACA results. There are two such cases in last-author ACA results and three in first-author results, one of which has three small factors. In other words, single-author-based ACAs appear to identify more detailed and more specialized areas of study within the more general and higher level ones that all-author ACAs show.

In addition, a small peripheral specialty in the all-author results (Signaling/Receptors) did not appear in the last-author results, and one in the other direction (an indistinct factor on cell senescence).

By contrast, four factors in the all-author results were missing in the first-author results (including a major one: Neuronal SCs and development), and three vice versa.

Some of the factors missing from the first-author map are fairly small on the other maps, and somewhat peripheral. Nevertheless, this provides evidence that first-author ACA provides a less comprehensive picture than the other maps.

The last-author map is the only one fragmenting one of the central generalist stem cell research areas, the Hematopoietic SCs specialty. This factor has many fairly low loadings on the other maps, but is always quite central, connecting to many other specialties across the major clusters. This specialty disintegrates into one that has close ties to the Embryonic SCs cluster and one with close ties to the Cancer cluster. The resulting subfactors are no longer identifiable as attempts to integrate the whole field.

It appears that single-author ACA tends to be more detailed, specialized and fragmented. As a result, there seem to be two types of research specialties that tend to be be different with single-author ACAs: (a) very small specialties (which disappear altogether when split up), and (b) large but diffuse, integrative research areas (which lose their integrative character when fragmented).

Findings from the present study seem to confirm some of those from previous studies, but do not prove the prediction that a higher level of collaboration in a research field would make the differences between first- and all-author ACAs more pronounced. We still see that all-author ACAs have a better model fit and favor highly successful groups of researchers in selecting researchers to include in the analysis. However, first-author ACA does not appear to identify more specialties or produce a less clear picture than all-author ACAs do. Instead, it appears that first-author ACA tends to show a more fragmented view of the field (including a separation of clusters into specialties), but can do so clearly and at a more detailed level. On the other hand, it sometimes does not capture the integration of the field as all-author ACAs do.

All in all, in the stem cell research field it appears that last-author ACA provides a much better approximation to all-author ACA than the traditional first-author ACA does. Since the last author listed in the byline of a paper is traditionally the head of the research lab that conducted the research reported in the paper (unless, of course, the head is the first author), the unit of analysis in last-author ACA is in effect the author's lab (or the author's success as the

leading scientist of a research lab) rather than the author as an individual. Last-author analysis thus provides a more comprehensive view of the researcher's research. On the other hand, just like first-author counting, last-author counting emphasizes a researcher's unique areas of study as the head of the lab whereas all-author counting evens these out with all of his or her contributions.

CONCLUSIONS

The present study compared all-author ACA with first- and last-author ACAs in the stem cell research field in terms of their effect on the mapping of the field, taking into account the highly collaborative nature and the unique collaboration and publishing culture of the field.

All three types of ACA are surprisingly similar in terms of the overall structure of the stem cell field they revealed, but do differ with respect to the degree of detail of major areas of studies shown and the number of specialties identified.

All-author ACA provides a more comprehensive picture of major specialties of the research field than the single-author-based ACAs do, and picks up some of the more subtle but potentially interesting aspects of the intellectual structure of the field. It offers a nice balance between showing the separation of major areas of studies and the integration of the entire field.

First-author ACA tends to emphasize a researcher's unique areas of study and most influential contributions. It therefore shows a considerably more detailed picture with more fragments within major clusters of specialties of the field than all-author ACA does.

Last-author ACA aggregates a researcher's contributions as a lab head, but still has the emphasis on individual's unique (as head of the lab) rather than all contributions. It therefore appears to be a compromise between first- and all-author ACAs in that it provides a slightly less comprehensive and more fragmented view of the field than all-author ACA does, but also a more comprehensive than first-author ACA does. However, it does not visualize the separation of major clusters of specialties as clearly as other types of ACA do.

All in all, all-author ACA still appears to be the winner although each type of ACA represents a different perspective of the field, each eliciting slightly different aspects of the intellectual structure of the field.

However, all-author ACA is much more complex and expensive to perform than first-author (and perhaps also last-author) ACA. As shown in the present study, it requires the combination of existing citation and other bibliographic databases through a multi-step process in order to collect the necessary data. On the other hand, first-author ACA is directly supported by both ISI and Scopus data, and last-author ACA (or an almost-all-author analysis in many research fields) is possible directly with Scopus data, which readily provides the names of last authors (even for papers with large numbers of authors).

Until citation databases provide better support for citation analysis studies in general and all-author counting in particular, it seems reasonable to use single-author-based ACA as a less expensive method producing reasonable approximations of all-author analysis results, given the relatively small differences between all-author and single-author-based ACAs and the expensive and complex process of data collection for a full all-author ACA. In this case, it is important to be aware that last-author ACA provides a much better approximation than the traditional first-author ACA does in research fields that have similar collaboration and publishing cultures as the stem cell research field. With Scopus, which claims and is shown to have excellent coverage of biomedical research, last-author ACA may well be a very efficient method for the study of biomedical fields and other fields where the same tradition holds and where core journal or keyword search based approaches to field delineation work.

ACKNOWLEDGEMENTS

This study was funded in part by the Social Sciences and Humanities Research Council (SSHRC) of Canada. The authors would like to thank Gencheng Guo for his assistance in the data collection process.

REFERENCES

Bordons, M., Gomez, I., Fernandes, M.T., Zulueta, M.A., & Mendez, A. (1996). Local, domestic and international scientific collaboration in biomedical research. *Scientometrics, 37*(2), 279-295.

Borgman, C.L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology, 36*, 3-72.

Brown, C. (2010). Communication in the sciences. *Annual Review of Information Science and Technology, 44*, 287-316.

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology, 52*(7), 558-569.

Department of Health and Human Services. (2001). *Stem Cells: Scientific Progress and Future Research Directions*. Retrieved May 21, 2010 from http://stemcells.nih.gov/info/scireport/execSum.asp.

Edge, D. O. (1979). Quantitative measures of communication in science: A critical review. *History of Science, 17*, 102-134.

Harter, S. P. & Kim, H. J. (1996). Electronic journals and scholarly communication: A citation and reference study. Proceedings of the Midyear Meeting of the American Society for Information Science, 1996, 299-315.

Garfield, E. (1979). *Citation Indexing — Its Theory and Application in Science, Technology, and Humanities*. New York: John Wiley & Sons.

Hair, J.F. Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate Data Analysis* (5th edition). Upper Saddle River, NJ: Prentice Hall.

Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science, 10*, 145-162.

Long, J. S., McGinnis, R., & Allison, P. D. (1980). The problem of junior-authored papers in constructing citation counts. *Social Studies of Science, 10*, 127-143.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: a critical review. *Journal of the American Society for Information Science, 40*, 342-349.

McCain, K. W. (1986). Cocited author mapping as a valid representation of intellectual structure. *Journal of the American Society for Information Science, 37*, 111-122.

Morris, S.A., & Van der Veer Martens, B. (2009). Mapping research specialties. *Annual Review of Information Science and Technology, 42*, 213-295.

Newman, M.E.J. (2001). The structure of scientific collaboration networks. *PNAS, 98*(2), 404-409.

Persson, O. (2001). All author citations versus first author citations. *Scientometrics, 50*(2), 339-344.

Schneider, J.W., Larsen, B., & Ingwersen, P. (2007). Comparative study between first and all-author cocitation analysis based on citation indexes generated from XML data. *Proceedings of the 11th International Conference on Scientometrics and Informetrics*, June 25-27, Madrid, Spain.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies, 4*, 17-40.

Smith, L. C. (1981). Citation analysis. *Library Trends, 30*, 83-106.

Sonnenwald, D.H. (2008). Scientific collaboration. *Annual Review of Information Science and Technology, 41*, 643-681).

Strotmann, A., Zhao, D., & Bubela, T. (2009). Author name disambiguation for collaboration network analysis and visualization. *Proceedings of The American Society for Information Science and Technology 2009 Annual Meeting*, November 6-11, 2009, Vancouver, British Columbia, Canada.

Strotmann, A., Zhao, D., & Bubela, T. (2010). Combining commercial and Open Access citation databases to delimit highly interdisciplinary research fields for citation analysis studies. *Journal of Informetrics, 4*(2), 194-200.

Sullivan, D., Koester, D., White, D. H., & Kern, R. (1980). Understanding rapid theoretical change in particle physics: A month-by-month co-citation analysis. *Scientometrics, 2*, 309-319.

Van Hooydonk, G. Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science, 48*, 944-945.

White, H. D. (1983). A cocitation map of the social indicators movement. *Journal of the American Society for Information Science, 34*, 307-312.

White, H. D. (1990). Author co-citation analysis: Overview and defense. In C. L. Borgman (ed.), *Scholarly communication and bibliometrics* (pp. 84-106). Newbury Park, CA: Sage.

White, H.D. (2003). Author cocitation analysis and Pearson's r. *Journal of the American Society for Information Science and Technology, 54*, 1250-1259.

White, H.D., & Griffith, B.C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science, 32*, 163-171.

White, H.D., & McCain, K.W. (1989). Bibliometrics. *Annual Review of Information Science and Technology, 24*, 119-286.

White, H.D., & McCain, K.W. (1997). Visualization of literatures. *Annual Review of Information Science and Technology, 32*, 99-168.

Zhao, D. (2005). Challenges of scholarly publications on the Web to the evaluation of science - A comparison of author visibility on the Web and in print journals. *Information Processing and Management, 41*(6): 1403-1418

Zhao, D. (2006a). Dispelling the myths behind straight citation counts. Information Realities: Shaping the Digital Future for All - *Proceedings of the American Society for Information Science and Technology 2006 Annual Meeting*, November 3 - 8, 2006, Austin, Texas, USA

Zhao, D. (2006b). Towards all-author co-citation analysis. *Information Processing and Management, 42*: 1578-1591.

Zhao, D., & Strotmann, A. (2008a). Comparing all-author and first-author co-citation analyses of Information Science. *Journal of Informetrics, 2*(3), 229-239

Zhao, D., & Strotmann, A. (2008b). Information Science during the first decade of the Web: An enriched author co-citation analysis. *Journal of the American Society for Information Science and Technology, 59*(6), 916-937.