



# Identifying Scientific Leadership in Stem Cell Biology

## A Systematic Approach

### Evaluating the Global Science Base

Peer reviewed publications are the hard output of academic science endeavor. In developing a view of fast moving academic research, analyzing the impact of scientific papers provides a mechanism for identifying the leaders in a particular field. BioLauncher recently undertook a systematic analysis of stem cell research to explore the characteristics of the academic science base, and were keen to develop an even handed methodology that was independent of our latent prejudices, or limited knowledge of the subject.

BioLauncher was able to systematically analyse the last decade of scientific publication information to identify the hot-spots of global research and the trends within it over the last 10 years. The process outlined below to generate this analysis is largely automated within the ShowcaseBio system

### Methodology

ShowcaseBio is a large semantically integrated database of information relating to life sciences. Amongst many other types of data, ShowcaseBio contains information useful for this particular analysis including:

- current and historic ISSN numbers (print and electronic) for all biomedical and life science journals
- journal impact factors and their trends over time<sup>1</sup>
- geocoding information for every address in PubMed, almost all research institutions and corporate research centres and every village, town and city worldwide

The methodology followed is shown below:

- All publications relating to the MeSH heading 'Stem Cells' for each calendar year period between 2000 and 2009 were identified from PubMed. Other more specific searches were also performed to narrow the scope of the analysis, but these data are not presented here.
- For each paper:
  - The authors' affiliation information was parsed and their location established
  - The journal that the paper was published in was identified and its Article Influence score for the year of publication was looked up
- Information was aggregated at a number of levels including by:
  - Researcher
  - Research institution or company
  - Geographic location (town/city)
  - Administrative area (e.g. US state, UK RDA region etc.)
  - Country and continent

---

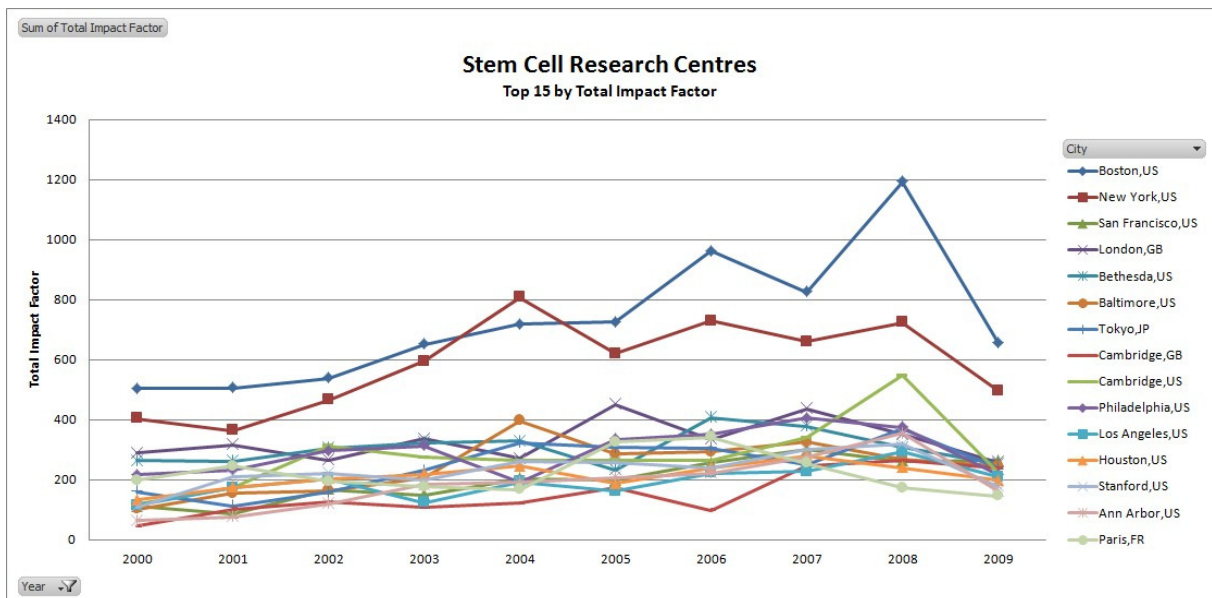
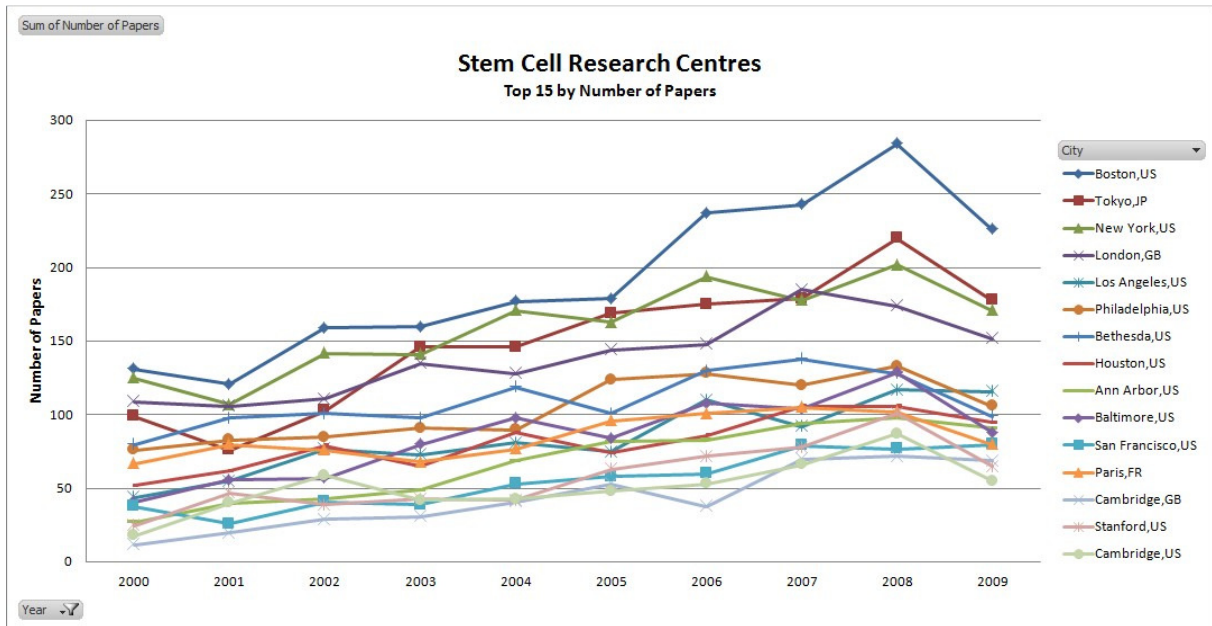
<sup>1</sup> Derived from Eigenfactor.org. Eigenfactor.org uses a sophisticated algorithm to ensure the most accurate reflection of the importance of each article. Eigenfactor.org uses full citation graph traversals, differences in citation frequencies, connections between disciplines and 5 year citation histories to generate its scores. For more information see <http://www.eigenfactor.org/methods.htm>

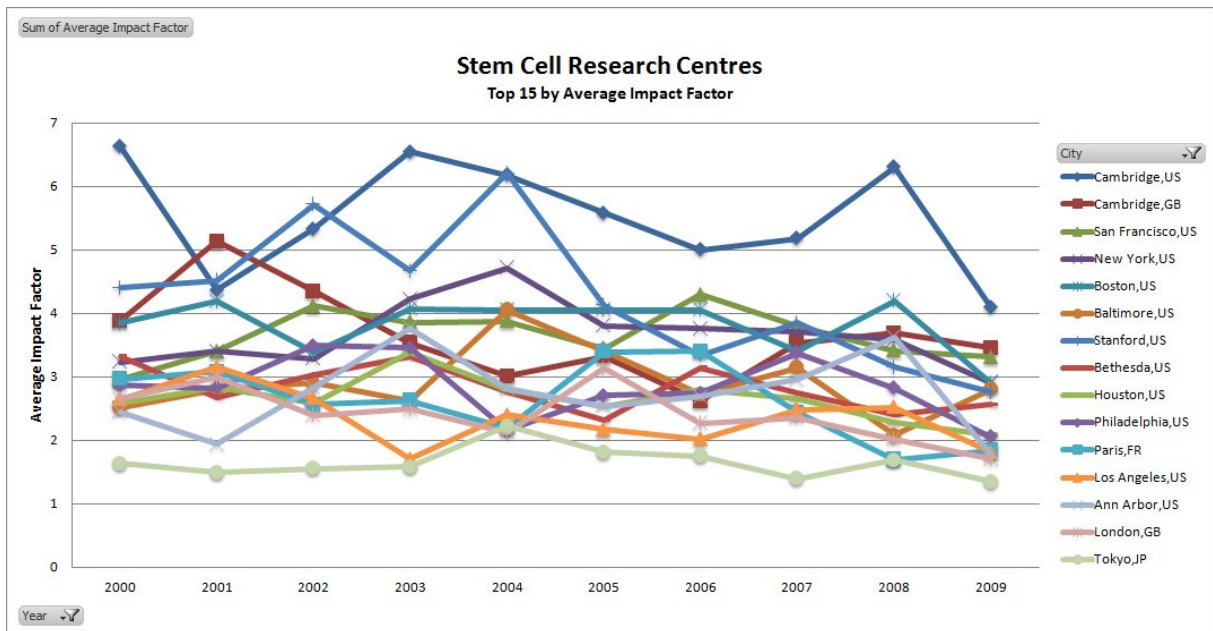
- For each date period and aggregation level, statistics were calculated including:
  - Number of papers published
  - Total (and average) impact factor
  - Data were normalized to account for the most recently published papers

## Results

The results presented here aggregate data based on the city of the last author's affiliation. This approach reveals the leading centers for stem cell research, without identifying individual researchers.

For each of the top 15 centres, BioLauncher investigated the total number of publications, the total number of publications by impact factor and finally investigated the average impact factor of publications. The summary findings are shown below.





## Conclusions

The number of publications graph provides insight into the research capacity of each city. Using this measure Boston, MA, Tokyo, Japan and New York, NY, have significant critical mass if stem cell investigators. However not all publications can be regarded as equally significant. Some publications are additive to the corpus of knowledge some are transformation in our understanding of stem cell biology. Considering the volume of publications together with the impact factor of each publication provides insight into the effectiveness of each research location.

Taking both publication quantity and impact factors into account, reveals that Boston, MA and New York, NY remain leading locations for stem cell biology. Tokyo, Japan slips in the rankings despite the ground breaking work of Yamanaka et al. in the field of induced pluripotent stem cells.

Thought leaders in a field are not defined by the capacity of their research. In fact it is plausible to advocate that major advances require more time and resources to achieve. If this is the case, these centers may be producing fewer publications. Regardless of the quantity of publication, we would expect that leaders in a field will produce papers that have a consistently higher average impact factor than the rest of the field.

The average impact factor analysis promotes Cambridge, MA and Cambridge, GB to the top of the rankings, suggesting that these centres though less productive in absolute numbers of publications published the highest impact science in the field.

These analyses are valuable tools in shaping the partnering strategies of companies that are seeking to engage with the global science base. If the publication data are aggregated against individual PIs instead of cities, the analysis can reveal the key opinion leader in a field. The method is not perfect (see below), but nevertheless offers a systematic approach to establishing a bench mark for engaging with important

## Discussion of the Method

This analysis represents one small example of the type of information that ShowcaseBio can generate. For more examples and information please see the ShowcaseBio webpage.

There are some limitations with the methodology outlined, including the following:

- Affiliation information is usually only associated with the first author of a paper. For collaborative papers with multiple authors at different centres this may mean that some attributions are missed. Since Pubmed often does not record multiple affiliation data, this is an intrinsic limitation of publication based analysis.

- All average impact factors are fundamentally based on citation metrics within a large but finite pool of publications, the majority of which are English language journals. This inevitably discriminates against papers that are written in languages other than English, and particularly against those using alternative alphabets. Whilst Eigenvector.org has tried to normalize some differences in citation frequencies (e.g. between domains), an element of this bias inevitably remains. Nonetheless, in practice since this bias is also inherent in the day to day process of scientific research, the analysis as presented remains an accurate summary of the likely influences on a scientist who is not fluent in French or Japanese (the two other languages represented in the Top 15 data presented).
- Not all addresses as provided in PubMed will geocode correctly and consequently cannot be added to a particular group. Badly formed or incomplete addresses may be excluded from this analysis.
- BioLauncher chose not to aggregate closely related place names, e.g. Stanford and Palo Alto or Cambridge and Boston, US. This is in part due to the highly subjective nature of such a condensation and partly because this analysis was one of a number performed at different levels of detail. BioLauncher has similar data by Researcher, Research Department, Research Institution/Company, Country and Continent.