

# Mapping of computational biology research inventions through quantitative patent analysis

Ankit Tripathi<sup>1,2</sup>, Tarakanta Jana<sup>2,3,\*</sup>, Nupur Wadia<sup>3</sup>, Deepak Bindal<sup>3</sup> and Swarnil Dey<sup>4</sup>

<sup>1</sup>Centre for Health Informatics, National Institute of Health and Family Welfare, Ministry of Health and Family Welfare, Government of India, New Delhi 110 067, India

<sup>2</sup>Academy of Scientific and Innovative Research, Mathura Road, New Delhi 110 025, India

<sup>3</sup>Patent Informatics Division, CSIR-National Institute of Science Communication and Information Resources, 14 Satsang Vihar Marg, New Delhi 110 067, India

<sup>4</sup>Rajiv Gandhi School of Intellectual Property Law, Indian Institute of Technology, Kharagpur 721 302, India

**Computational biology is relatively a young branch of science, which has experienced tremendous growth in the last two decades. The seeds of this inter-discipline were sown at the end of the seventies, when computers became simpler to use, some biology laboratories decided to adopt them, mainly for storing and managing genomic data. As a result, there was quick completion of projects that would have otherwise taken several years. With a snowballing effect, genome sequencing projects generated large amounts of data whose management required more powerful computers. The advent of the internet has allowed all the research laboratories to share their data, and make them available worldwide through some new genomic and proteomic data banks, such as Gene-Bank, European Molecular Biology Laboratory Data Library and many more. A patent portfolio shows the state of the art of research and development in a specific discipline. Patenting an invention is still tough in many countries, especially when the major part of the claim invention involves mathematical formulae, algorithms and computer programs. These per se are not patentable in many countries; however, they can be protected under copyrights as literary works or trade secrets. It is well known that mathematical formulae, algorithms and computer programs mainly form the core of computational biology. Till date, no study has been made to measure inventive progress in this new inter-discipline. The study shows the research and invention have occurred in a wide range of inter-disciplines, including functional genomics, molecular structure, sequence comparison, molecular simulations, machine learning, data visualization, database development and inventive activity confined to a handful of industrially developed and developing nations, of which majority are privately owned.**

**Keywords:** Bioinformatics, computational biology, patent analysis, research trends inventions.

COMPUTATIONAL biology has transformed conventional laboratory-based biotechnology into a computer-based

science that focuses on automated collection, compilation, storage, retrieval and analysis of biological data<sup>1</sup>. The history of computational biology dates back to the 1960s, when Margaret Dayhoff analysed a large set of protein sequences to study molecular evolution<sup>2</sup>. In India, the history of bioinformatics can be traced back to the 1960s, when G. N. Ramachandran and his colleagues derived the famous Ramachandran plot that provided the foundation of modern structural biology/bioinformatics<sup>3,4</sup>. In the 1970s, the publication of Needleman and Wunsch<sup>5</sup> on sequence analysis acted as the starting point for the development of numerous follow-up sequence alignment algorithms. In the 1990s, large amounts of experimental data were produced by sequencing projects, and the development of databases and algorithms for sequence alignment attracted attention to obtain further knowledge from genomic and protein sequences. Since then there has been a rapid growth in the number of full genome sequences, which are freely available to the public in GeneBank. Similarly, a large number of protein sequences are available in the Protein Data Bank. To explore knowledge from the genomic and protein sequences, several computational biology software tools have been developed since the late 1990s (ref. 6). Computational biology and bioinformatics, the terms often used interchangeably, represent a rapidly evolving biological discipline. It is a broad discipline that includes: (i) the management, analysis and integration of diverse types of biological data; (ii) the modelling of biological systems and biological structures, and (iii) the use of computational techniques to support and enable virtually all areas of modern life science results. The field is undergoing rapid evolution and growth, and will continue to expand in its scope in the years to come.

Patenting an invention is still difficult in many countries, especially when the major part of the claim invention involves mathematical formulae, algorithms and computer programs, which form the core of bioinformatics. Intellectual property (IP) associated with bioinformatics and computational biology has many dimensions, such as lines of code, algorithms, data content and structures, and user interfaces. There are several ways in

\*For correspondence. (e-mail: tkj@niscair.res.in)

## RESEARCH ARTICLES

**Table 1.** International Patent Classification (IPC) codes (effective from 1 January 2011) for searching/accessing bioinformatics and computational biology patents

IPC search codes	Description of subject
G06F 19/10	Bioinformatics, i.e. methods or systems for genetic or protein-related data processing in computational molecular biology ( <i>in silico</i> methods of screening virtual chemical libraries; <i>in silico</i> or mathematical methods of creating virtual chemical libraries).
G06F 19/12	For modelling or simulation in systems biology, e.g. probabilistic or dynamic models, gene-regulatory networks, protein interaction networks or metabolic networks.
G06F 19/14	For phylogeny or evolution, e.g. determination of evolutionarily conserved regions or phylogenetic tree construction.
G06F 19/16	For molecular structure, e.g. structure alignment, structural or functional relations, protein folding, domain topologies, drug targeting using structure data, involving two-dimensional or three-dimensional structures.
G06F 19/18	For functional genomics or proteomics, e.g. genotype–phenotype associations, linkage disequilibrium, population genetics, binding-site identification, mutagenesis, genotyping or genome annotation, protein–protein interactions or protein–nucleic acid interactions.
G06F 19/20	For hybridization or gene expression, e.g. microarrays, sequencing by hybridization, normalization, profiling, noise correction models, expression ratio estimation, probe design or probe optimization.
G06F 19/22	For sequence comparison involving nucleotides or amino acids, e.g. homology search, motif or single nucleotide polymorphism discovery or sequence alignment.
G06F 19/24	For machine learning, data mining or biostatistics, e.g. pattern finding, knowledge discovery, rule extraction, correlation, clustering or classification.
G06F 19/26	For data visualization, e.g. graphics generation, display of maps or networks or other visual representations.
G06F 19/28	For programming tools or database systems, e.g. ontologies, heterogeneous data integration, data warehousing or computing architectures.

which bioinformatics and computational biology IP can potentially be protected – patent, copyright and trade secret<sup>7</sup>. Among these, patent protection is important in protecting IP associated with bioinformatics and computational biology. Data content is not protectable by patents. Copyright offers a way of protecting lines of code and the user interface, but tends to be thin. It protects unauthorized copying, modification or distribution, not independent development, and the trade secret route may help in protecting technical knowhow. Using approaches developed in the fields of information technology and business, patent applicants now seek to protect certain facets of their inventions, which include software, methods of doing business and uses of information as well as more traditional biotechnological products and processes<sup>8</sup>. Though analysis of informatics on bioinformatics is challenging because of its fast evolving nature and close transition from technique (and technology)-driven science to an information-driven science<sup>9,10</sup>, yet subtle efforts have been made to map research trends in bioinformatics using published literature from the National Centre for Biological Information (NCBI), PubMed and Web of Science<sup>11–13</sup>. Access of bioinformatics patents has limitation due to lack of proper search strategy<sup>14,15</sup>. To the best of our knowledge, there are no previous reports where global patent data have been collected based on the new International Patent Classification (IPC) systems. The present study provides information regarding annual distribution, technological growth, geographical preference and different star assignees (applicants) in this fast-evolving discipline and its knowledge spillovers.

### Methodology

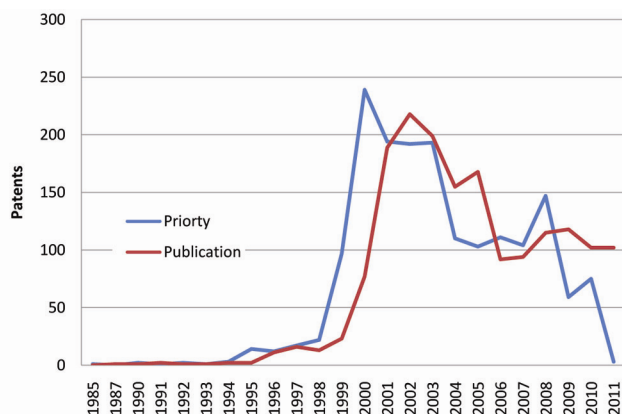
The search methodology for obtaining patent publications relating to computational biology was based on IPC searches. World Intellectual Property Organization (WIPO) introduced new alpha-numerical search codes to cover computational biology patents in the new IPC system which was effective from 2011. Table 1 shows IPC codes for inter-disciplinary subject areas for searching bioinformatics and computational biology patents. Here, analysis has been done for all issued patents from 1985 to 2011. There were about 7511 patent documents retrieved from Thomson Innovation database during the period which on analysis resulted in 7430 patent documents, that were further analysed on basis of International Patent Documentation Centre (INPADOC) patent family. This gave 1702 INPADOC patent families, which were analysed in this study.

### Annual distributions for understanding evolving research trends and life of patents

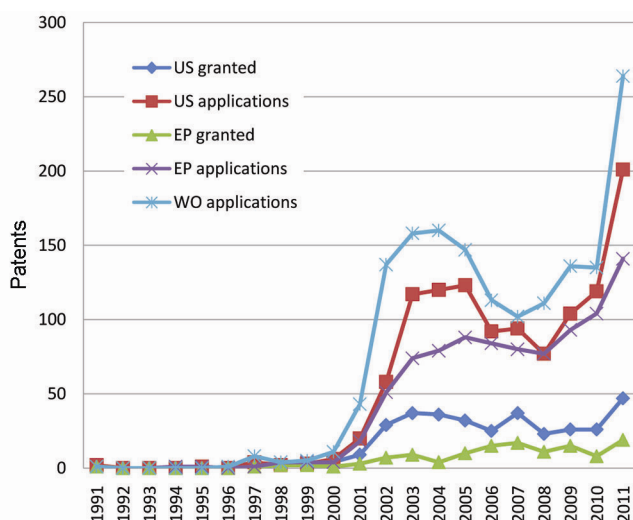
In general, a patent term is 20 years. However, technology and services in computational biology are generally short-lived and normally have economic yield only for a limited period of time even if the life of such patents is normally for 20 years. The most common feature of such an arrangement is that equipment, software and data are generally reclaimed by the owner upon termination of the agreement. The computational biology services also usually come with milestones and royalties attached to services

that are made for actual research programmes. The computational biology service model of the mid-1990s is no longer attractive to investors in the present-day context due to rapid innovation and obsolescence of software and services. So companies have started to upgrade their technologies and services to stay in the competition. In this study, we take a close look at the growth of patents in computational biology from 1985 to 2011. An understanding of annual patenting activity is necessary to determine the life of a patent and growth of research and invention over time in this inter-discipline. The earliest patenting activity in computational biology was seen in the early 90s. Till 1998, there was no significant growth in patenting activity in this inter-discipline. From 1999 to 2003, there was significant rise in patent applications, but thereafter a decline was observed from 2004 to 2009 (Figure 1).

This could be due to the fact that breakthrough in patenting activity came when the Federal Circuit Court of



**Figure 1.** Annual distribution of patents on computational biology (based on priority filing and publication year).



**Figure 2.** Annual distribution of computational biology patents (based on applications and granted patents).

Appeals in case of *State Street Bank & Trust Co. v. Signature Financial Group, Inc.* in 1998, broke barriers for patenting computer-based business models, as the case explicitly struck down the longstanding ‘business method’ exception and severely limited the mathematical algorithm exception in patent law of USA<sup>16</sup>. Earlier to this liberal revolutionary judgment, the mathematical algorithm exception generally prevented computer programs that consisted of a mathematical algorithm from being patented. Implementation of this judgment revolutionized computational biology patenting activity in 1998 in USA and had an impact on research, development and protection globally.

Figure 2 shows that a total 1143 patents were filed in USA, out of which 340 were granted. On the other hand, European Patent Office (EPO or EP) granted only 125 out of 905 patent applications. The present study shows the liberal view of USA for granting patents on computational biology compared to Europe, and also the huge growth of patents in computational biology from 2000 to 2011.

### Growth of inventions in prominent sub-disciplines of computational biology

In order to understand the extent to which the new inter-disciplines are growing, the collected data have been segregated and studied. The study shows there are few prominent research areas where maximum inventive activity has occurred in order to solve emerging problems in modern biology where computational interventions are of paramount importance. Prominent research areas are functional genomics or proteomics (G06F 19/18), molecular structure (G06F 19/16) and sequence comparison (nucleotides or amino acids) (G06F 19/22), machine learning, data mining and biostatistics (G06F 19/24), hybridization and gene expression (G06F 19/20), programming tools and database systems (G06F 19/28), modelling or simulations in system biology (G06F 19/12), data visualization (G06F 19/26), etc. For instance, functional genomics or proteomics had maximum (443 patent applications) followed by molecular structure (416 patents), sequence comparison (376 patents), machine learning, data mining and biostatistics (341 patents), hybridization and gene expression (334 patents), programming tools and database systems (280 patents), modelling or simulations in system biology (146 patents) and data visualization (143 patents; Figure 3).

### Geographical preference for navigating greater market access

The global locus of patenting activity in computational biology has been changing dramatically from the first decade of the 21st Century. Applicants and inventors

generally prefer a country for filing patents, where they can protect their inventions to recoup their investments through monetization. Figure 4 shows 18 patent offices and 2 regional routes (Patent Cooperation Treaty, i.e. WIPO or WO, and European Patent Office, i.e. EPO or EP), where maximum patents on computational biology are filed. The study shows that WIPO or WO route is the most active channel for the filing of patents. Maximum number of patents is filed in USA followed by Australia (AU), Canada (CA), Japan (JP), China (CN), Germany (DE), Great Britain (GB), Austria (AT), Israel (IL) and South Korea (SK; Figure 4). EPO or EP is also a favourite choice, as it ranks third in patent filing. Among Asian countries, Japan, China and South Korea show positive growth of patent filing on computational biology. Our study reveals that global patenting activities are confined to a handful of industrially developed and developing countries.

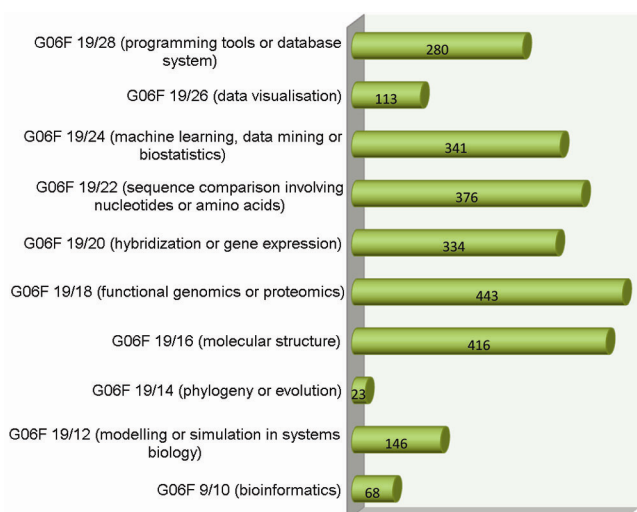


Figure 3. Research invention growth in various sub-disciplines of computational biology.

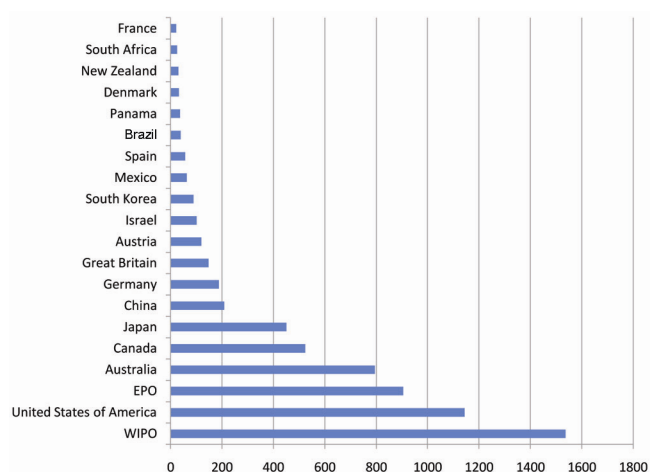


Figure 4. Geographical distribution of computational biology patents.

The present study corroborates the study of Grand View Research, Inc<sup>17</sup>, which suggests that North America is the largest regional market in computational biology, accounting for 58.0% of the global revenue in 2013. Its large share is majorly attributed to the growing research investments and initiatives for development of drug discovery, disease modelling technologies and innovations in biological computation methods. It is expected that Asia Pacific will witness the fastest growth in the market, at an estimated cumulative annual growth rate (CAGR) of 28.0%, owing to the increasing expenditure in clinical studies for pharmacokinetics and pharmacogenomics.

The application of computational biology in proteomics for cellular simulation is expected to grow at the fastest CAGR of over 22% owing to the shift in sequencing studies towards further analysis of the nature of various biological proteins as a part of development of personalized medicine. The Grand View Research report suggests that drug discovery and disease modelling applications dominated the market, accounting for 58.0% of global revenue in 2013. Computational biology databases dominated the tools segment and accounted for 44.5% of the market revenue in 2013, owing to extensive genomic database usage for warehousing and data mining of biological and sequencing data generated from genomic and proteomic studies. Simultaneously, the software and services for the analysis of biological data are expected to witness the fastest CAGR of over 21.0%.

### Public and private sector contribution for R&D growth

There are star applicants who have contributed to R&D growth in these rapidly evolving inter-disciplines. Figure 5 shows the top 20 assignees/applicants who contributed greatly in research and protection of inventions in computational biology. Our analysis shows that majority of

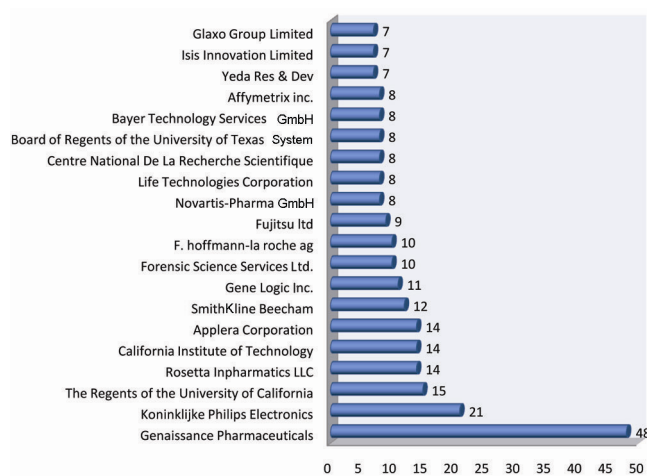


Figure 5. Top assignees/applicants in computational biology patents.

those featuring in the top 20 list is privately owned. For instance, University of California, University of Texas, California Institute of Technology and Centre National de la Recherche Scientifique are the only public funded institutes appearing in the top 20 applicants list. Genaiasance Pharmaceuticals, with 48 patent publications, filed the maximum number of patents. This is followed by Koninklijke Philips Electronics and University of California with 21 and 15 respectively. We have also observed that most of the assignees are located in USA and Europe. While funding from the public sector continues to fall, when public needs and scientific complexity are on the rise, private sector collaboration with academia and government has become a key in furthering research in this field. For instance, earlier studies have shown that public sector research institutions have performed the upstream, basic research to elucidate the underlying mechanisms and pathways of disease and identify promising points of intervention, whereas corporate researchers have performed the downstream, applied research to discover drugs that can be used for treating diseases, followed by patenting the same and follow up the development activities to bring the drugs to market<sup>18,19</sup>. Recently, quantitative analysis has revealed that the United States National Institutes of Health (NIH) funding increases total private-sector patenting<sup>20</sup>. The study suggests that an additional USD 10 million in NIH funding for research in pharmaceuticals and healthcare generates 3.26 additional private-sector patents in that area, or roughly one patent for every two NIH grants.

The commercialization of computational biology has spawned a support industry that provides bioinformatics tools, a function previously provided by university researchers. The companies are providing new data analysis tools and software platforms for data management, expression profile analysis, links to sequence and annotation databases, function prediction based on pathway information, data mining and data tracking of automated processes. Companies that implement these new bioinformatics tools and software platforms are faced with problems that arise when trying to compare, store and analyse data produced from multiple platforms<sup>21</sup>.

The global market for computational biology is expected to reach USD 4285.1 million by 2020 growing at a CAGR of 21.1%, according to a study by Grand View Research, Inc.<sup>17</sup>. Another market report published by Transparency Market Research<sup>22</sup> predicted that the computational biology market will rise at a CAGR of 21.3% from 2012 to 2018 and reach an estimated value of USD 2937 million by 2018. Steady increase in the usage and application of computational biology for bioinformatics R&D programmes designed for sequencing genomes to better understand biological systems and increasing the number of clinical studies in pharmacogenomics and pharmacokinetics for novel drug discovery studies, is expected to drive the computational biology demand over

the next five years. Computational genomics for cellular and biological simulation applications dominated the market in 2013, accounting for over 47% of the global revenue. Key factors driving demand for these applications include the growth of drug designing, disease modelling and personalized medicine applications, and the growing usage of computational biology for functional and structural genomics, epigenomics and metagenomics for the analysis of protein transcription, protein–protein interactions, gene sequencing and expression, and three-dimensional protein structure analysis.

### Technology diffusion and knowledge spillover

Patent citation data are used in a growing body of economics and business research to determine technological diffusion and knowledge spillover, though the tools have been misused in many cases<sup>23–25</sup>. Generally, strength of patents is measured by the number of citations made by other patents that are issued later<sup>26</sup>. The belief that such knowledge spillover is an important mechanism for growth productivity has been a feature of policy debates since the end of World War II and has also figured prominently in economic scholarship on technological change<sup>27,28</sup>. We studied this claim directly by identifying the number of citations in computational biology that explicitly cite other patents issued later. Table 2 shows the dominant patents on the basis of citations. International Business Machine, USA has the most cited patents in bioinformatics (US5418944A), which concern molecular retrieval system and method using a hierarchy of molecular structures. The second most highly cited patent application, US20040093331A1, assigned to University of Texas claims a method for data mining system in the medical field, and includes a knowledge discovery engine in which relationships between integrated objects are identified, retrieved, grouped, ranked, filtered and numerically evaluated. Another patent application, US20050187916A1 assigned to Eugene Levin and Martha Elizabeth Corey, also claims a method for sequential data comparison in bioinformatics, and involves generating indication of match of target and query sequences based on ordered comparison of binary strings of query sequence with corresponding strings of target sequence. The remaining inventions in Table 2 are focused on developing computational method for analysing biological data, application of computational methods, statistics, mathematical modelling, algorithms and database structure.

### Conclusion

Computational methods have become indispensable to modern biology, especially for biomedical and pharmaceutical industry in order to acquire new knowledge or

## RESEARCH ARTICLES

**Table 2.** Top cited patents on bioinformatics and computational biology

Publication no.	Title	Publication date	Assignee/ applicant	Inventor	Citing of patents
US5418944A	Knowledge-based molecular retrieval system and method using a hierarchy of molecular structures in the knowledge base	1995-05-23	International Business Machines Corporation, USA	Dipace, Luigi and Fabrocini, Filippo	58
US20040093331A1	Computer program products, systems and methods for information discovery and relational analyses	2004-05-13	University of Texas System, USA	Garner, Harold, R. Wren and Jonathan, D.	45
US20050187916A1	System and method for pattern recognition in sequential data	2005-08-25	Levin, Eugene and Corey, Martha, USA	Levin, Eugene and Corey, Martha	37
WO1999028505A1	Methods and devices for measuring differential gene expression	1999-06-10	Curagen Corporation, USA	Rothberg, Jonathan, M. Nallur, Girish, N. Hu, Xinghua	27
US20050165594A1	System, method and apparatus for causal implication analysis in biological networks	2005-07-28	Genstruct Inc., Cambridge, USA	Chandra, Dundee Chandra, Maria Segaran, Suresh Kightley, David Sun, Justin Pratt, Dexter	22
US5598350A	Genetic motif extracting method and apparatus	1997-01-28	Fujitsu Limited, Kanagawa, Japan National Institute of Genetics, Shizuoka, Japan	Kawanishi, Yuichi Gojobori, Takashi Tateno, Yoshio Ikeo, Kazuho Kawai, and Masahito	21
WO2001031317A1	Method and apparatus for selectively retrieving biological samples for processing	2001-05-03	Genometrix Genomics Incorporated, USA	Hogan, Michael, E. Brignac, Stafford Jr, J. and King, Terri	18
US5642292A	Methods for searching stable docking models of biopolymer–ligand molecule complex	1997-06-24	Itai Akiko, Tokyo, Japan	Itai, Akiko and Mizutani, Miho	18
WO2001069430A1	Database system and method	2001-09-20	DNA Sciences Inc., USA	Rienhoff, Y., Hugh, Jr. Kean, James, R. Jones, and Hywel, B.	16
WO2002025519A1	Gene diagnosis information providing method, diagnosis information providing terminal, and diagnosis information receiving terminal	2002-03-28	Kabushiki Kaisha Toshiba, Japan	Takada, Yoichi Suzuki, Yoshinori	15
US20020087275A1	Visualization and manipulation of biomolecular relationships using graph operators	2002-07-04	Kim, Junhyong & Jiang, Shan, USA	Kim, Junhyong and Jiang, Shan	15
WO2001030808A1	Methods and compounds for modulating melanocortin receptor–ligand binding	2001-05-03	University of California, USA	Millhauser, Glenn, L. Bolin, Kimberly, A. Anderson and D., Joe	15
WO2001020043A1	Method of cluster analysis of gene expression profiles	2001-03-22	Affymetrix Inc., USA	Hu, Jing-shan Durst, Mark Khurgin, Elina Balban and David, J.	15
WO2000029987A1	Methods for identifying and classifying organisms by mass spectrometry and database searching	2000-05-25	University of Maryland, USA	Demirev, Plamen, A. Fenseleau, Catherine	15

(Contd)

Table 2. (Contd)

Publication no.	Title	Publication date	Assignee/ applicant	Inventor	Citing of patents
US6029114A	Molecular modelling of neurotrophin-receptor binding	2000-02-22	Queen's University at Kingston, USA	Shamovsky, Igor L. Ross, Gregory M. Riopelle, Richard J. Weaver and Donald F.	15
US20030187587A1	Database	2003-10-02	Swindells, Mark Thornton, Janet and Jones, David, Great Britain	Swindells, Mark Thornton, Janet and Jones, David	14
WO1997014106A1	Identification of common chemical activity through comparison of substructural fragments	1997-04-17	Terrapin Technologies Inc., USA	Villar, Hugo, O. Bone and Richard, G. A.	14
US4982338A	Method for processing information on chemical reactions	1991-01-01	Fuji Photo Film Co Ltd, Minami Ahsigara, Japan	Fujita, Shinsaku	14
US20030165952A1	Method and an algorithm for mRNA expression analysis	2003-09-04	Linnarsson, Sten Ernfor, Patrik and Bauren, Goran, Sweden	Linnarsson, Sten Ernfor, Patrik Bauren, Goran	13
WO1999053101A1	Identification of genetic targets for modulation by oligonucleotides and generation of oligonucleotides for gene modulation	1999-10-21	Isis Pharmaceuticals Inc., USA	Cowsert, Lex, M. Baker, Brenda, F. Mcneil, John Freier, Susan, M. Sasmor, Henri, M. Brooks, Douglas, G. Ohasi, Cara Wyatt, Jacqueline, R. Borchers, Alexander, H. Vickers, Timothy, A.	13

use of existing knowledge in a creative manner so as to generate new concepts and methodologies. The present study indicates that there has been unprecedented growth in the number of patent filings in the first decade of the 21st century. The study also reveals that USA continues to lead innovation in computational biology as reflected in the increased patent filing. This is possibly because of the mobilization of resources, nurturing the passion of R&D personnel and market force in computational biology and liberal patent laws at the policy level. The other countries which follow liberal policies are Australia, Canada, Japan, China, Germany, Great Britain, Austria, Israel and South Korea. The global shift is interesting as it shows important market destination for protecting inventions and to return the investments in this newly emerged information- and knowledge-driven discipline. In the present study, we have observed that filing of patents in computational biology is low in most of the countries. We presume that each country has its own set of circumstances surrounding its R&D objectives – so there is no ‘one-size-fits-all’ solutions that is applicable across countries and technologies. The study shows that the WIPO route is the most preferred for applying patents, as it is cost-effective. R&D for computational biology has moved into the fourth paradigm, facilitating data-intensive discoveries in biological science. Our study

shows that blending of computational biology has occurred on a wide range of subject areas, including structural biology, gene expression, molecular simulations, machine learning, data visualization, database development, etc. The study also shows top applicants who have contributed considerably to the fast growth of this discipline and knowledge spillover of important patents.

- Hultquist, S. J., Harrison, R. and Yang, Y., Patenting bioinformatic inventions: emerging trends in the United States. *Nature Biotechnol.*, 2002, **20**, 743–744.
- Mount, D. W., *Bioinformatics: Sequence and Genome Analysis (Second Edition)*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2004, p. 8.
- Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V., Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 1963, **7**, 95–99.
- Kulkarni-Kale, U., Sawant, S. and Chavan, V., Bioinformatics education in India. *Brief. Bioinform.*, 2010, **2**, 616–625.
- Needleman, S. B. and Wunsch, C. D., A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 1970, **48**, 443–453.
- Altschul, S. et al., The anatomy of successful computational biology software. *Nature Biotechnol.*, 2013, **31**, 894–897.
- McBride, M. S., Bioinformatics and intellectual property protection. *Berkley Technol. Law J.*, 2002, **17**.
- Maschio, T. and Kowalski, T., Bioinformatics – a patenting view. *Trends Biotechnol.*, 2001, **19**, 334–339.

## RESEARCH ARTICLES

---

9. MacMullen, W. J. and Denn, S. O., Information problems in molecular biology and bioinformatics. *J. Am. Soc. Inf. Sci. Technol.*, 2005, **56**, 447–456.
10. Barlett, J. C. and Toms, E. G., Developing a protocol for bioinformatics analysis: an integrated behavior and task analysis approach. *J. Am. Soc. Inf. Sci. Technol.*, 2005, **56**, 469–482.
11. Patra, S. K. and Mishra, S., Bibliometric study of bioinformatics literature. *Scientometrics*, 2006, **67**, 477–489.
12. Manoharan, A. *et al.*, Bioinformatics research – an informetric view. In International Conference on Information Communication and Management, Singapore, IACSIT Press, 2011, 199–204.
13. Song, M. *et al.*, Productivity and influence in bioinformatics: a bibliometric analysis using PubMed Central. *J. Am. Soc. Inf. Sci. Technol.*, 2014, **65**, 352–371.
14. Rasmussen, B., Commercialization processes in bioinformatics: analysis of bioinformatics patents. CSES Working Paper No. 26, Centre for Strategic Economic Studies, Victoria University, Melbourne, 2005.
15. Park, H. S., Preliminary study of bioinformatics patents and their classification registered in the KIPRI database. *Genomics Inf.*, 2012, **10**, 271–274.
16. Rockman, H. B. (ed.), *Protection of Computer Related Invention. Intellectual Property Law for Engineers and Scientist*, John Wiley, 2004.
17. Computational biology market analysis by application (cellular & biological simulation, pharmacogenomics, drug discovery, drug development, lead optimization, lead discovery, pharmacokinetics, pharmacodynamics, disease modelling, clinical trials), by service (in-house, contract), by end-user (academics, industry, commercial) and segment forecasts to 2002; <http://www.grandview-research.com/industry-analysis/computational-biology-market>
18. Cockburn, I. M. and Henderson, R. M., Public-private interaction in pharmaceutical research. *Proc. Natl. Acad. Sci. USA*, 1996, **93**(23), 12725–12730.
19. Cockburn, I. M. and Henderson, R. M., Publicly funded science and the productivity of the pharmaceutical industry. *Innovation Policy Econ.*, 2000, **1**, 1–34.
20. Azoulay P., Graff Zivin, J. S., Li, D. and Sampat, B. N., Public R&D investments and private-sector patenting: evidence from NIH funding rules. NBER Working Paper 2015; <http://www.nber.org/papers/w20889.pdf>
21. Global bioinformatics industry – market research report, 2014; <http://www.reportlinker.com/p092468/World-Bioinformatics-Market.html>
22. Computational biology market (pharmacodynamics, cellular modelling, computational genomics, proteomics, pharmacogenomics, pharmacokinetics, human simulation software, drug discovery and development) – global industry analysis, size, share, growth, trends and forecast, 2012 – 2018; <http://www.transparencymarket-research.com/computational-biology.html>
23. Finkel, A., Powering up citations. *Nature*, 2014, **511**, S77.
24. Adams, J., The citation game. *Nature*, 2014, **510**, 470.
25. Cronin, B. and Sugimoto, C. R. (eds). *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, MIT Press, 2014.
26. Karki, M. M. S., Patent citation analysis: a policy analysis tool. *World Patent Inf.*, 1997, **19**, 269–272.
27. Nelson, R. R., The role of knowledge in R&D efficiency. *Q. J. Econ.*, 1982, **97**(3), 453–470.
28. Cockburn, I. M. and Henderson R. M., Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *J. Ind. Econ.*, 1998, **46**(2), 157–182.

Received 18 December 2014; revised accepted 5 December 2015

doi: 10.18520/cs/v110/i7/1307-1314

---