Genomics and IP: An Overview

Anandita Singh, Sandip Das & Neeti Wilson

TERI School of Advanced Studies, Darbari Seth Block, Habitat Center, Lodhi Road, New Delhi 110 003, India
Department of Biotechnology, Hamdard University (Jamia Hamdard), Delhi 110 062, India
Anand and Anand, K 47 Kailash Colony, New Delhi 110 048, India

Received 11 October 2006

The field of genomics and systems biology promises to provide unconventional solutions to problems relating to healthcare and agriculture. The various biotechnologies centred on genomics are undeniably a dominant force in the world of economics and the stakes are enormous. The field of genomics and system biology is immensely cross-sectoral and flourishes the most when the knowledge from various domains are integrated. However, this very essence of genomics adds layers of complexities while segregating the contributions from various domains related to IPRs. The foremost technologies guiding the advancements in the field of genomics include genome-wide sequencing, high throughput expression profiling, bioinformatics and the resultant databases.

The need to protect and capitalize innovation is sacrosanct to best business practices. At the same time, providing access to the benefits accruing from such innovations to public is of importance to the society at large. IPR regime attempts to meet these seemingly conflicting objectives of providing incentive to the innovators and yet acting as a disclosure mechanism for promoting a continuum in research and development. This dual reward-and-disclosure clause is beneficial to both the inventor and the end user. This article attempts to map the high throughput genomic technologies along with their attendant databases and analysis tools. Through this outline, the article makes an attempt to inform the extent of IP protection and issues emanating from the application of such technologies.

Keywords: Genomics, IP protection, ESTs, microarrays, DNA sequencing

Genomics is the application of high throughput tools and technologies at a genome-wide scale to generate biological data pertaining to knowledge on genome and its expression state. The storage, management and analysis of this information relies on computational and statistical tools. In essence, genomics involves integration of datasets from every conceivable knowledge portal in order to address a definite biological problem.

Quintessentially, the key technologies of genomics include high-throughput sequencing; microarrays for analysing global expression profiling, re-sequencing for screening and detection of genome wide polymorphisms, alternative splicing, methylation and DNA binding states. An important outcome of genomics relates to defining the function of genomic sequences.

Outstanding advancements in the field of genomics have taken place predominantly in the developed countries, mainly the USA and European nations, as evidenced by its commercial application. These countries therefore are forerunners in evolving an IPR regime in respect of biotechnology and genomics. In contrast, developing countries like India initially lagged behind in the area of genomics related to IPRs. The two main reasons for slow start of India in the genomic scenario may be attributed to:

- India is technologically not at par with the developed nations as far as genomics is concerned;
- The Indian Patent protection and enforcement system was not very strong until recently, resulting in the inventions emanating from Indian R&D laboratories to be preferably directed to the US and/or EP for patent protection.

For any invention to be awarded a protection, following criteria must be met as per the guidelines:

- Appropriate subject matter
- Novelty
- Non-obviousness
- Utility
Apart from these, a patent application must contain written description including specifications that disclose the best mode of making the invention work. Genomic discoveries are an outcome of significant human interventions and thus are actually inventions, which are eligible for patent protection provided that the aforementioned patent doctrines are met with.

It is worth mentioning at this stage that the patent system of the US being most peculiar than rest of the world, anything under the sun maybe patented so long it is manmade. The practice of Europe in terms of granting protection to DNA sequence related inventions is evident from Box 1 wherein the European directives have been provided.

This article explores the field of genomics and related IP issues in context of US and EU practices, which have specialized guidelines. An overview of the developing IP scenario, particularly, relevant to genomics, in India has been briefly discussed in the end.

The genomic workspace is a complex landscape wherein massive amount of biological information is generated, curated, annotated and managed as structured databases. An integral feature of genomics is maintaining coordinated exchange of information amongst these databases. This simultaneous flow of information in multiple directions makes the entire field highly dynamic. Extensive applications of bioinformatics tools are hence imperative. Integration of datasets from various sources adds complexities in data interpretation. For the sake of simplicity, we have split the genomics landscape into four sections and analysed the IPR issues in all its ramifications in a stepwise fashion.

**Genomic Technologies**

**High Throughput Sequencing**

The enormous progress in sequencing technology including the drastic reduction in sequencing cost⁵ has led to an explosion in the amount of data that is generated. At present count there are nearly 250 higher eukaryotes that are being sequenced using public funding⁶, not including the thousands of microbes, and privately funded sequencing efforts. The projected sales for the worldwide DNA sequencing and proteomics markets are expected to rise at an average annual growth rate of 17.6% from US $7.8 billion in 2004 to US $17.5 billion in 2009.⁷

A patent search in USPTO with key words selected as ‘DNA sequencing’ in title, revealed close to 180 hits. Typically, the IPRs related to sequencing technology fall within the following categories:

- Sequencing methods and dye chemistries
- Sequence per se (genomic, EST, GSS)

**Sequencing Methods and Dye Chemistries**

The world sequencing market is dominated by Applied BioSystems (ABI) which uses the BigDye®

---

**BOX 1**


(22) Whereas the discussion on the patentability of sequences or partial sequences of genes is controversial; whereas, according to this Directive, the granting of a patent for inventions which concern such sequences or partial sequences should be subject to the same criteria of patentability as in all other areas of technology: novelty, inventive step and industrial application; whereas the industrial application of a sequence or partial sequence must be disclosed in the patent application as filed;

(23) Whereas a mere DNA sequence without indication of a function does not contain any technical information and is therefore not a patentable invention;

(24) Whereas, in order to comply with the industrial application criterion it is necessary in cases where a sequence or partial sequence of a gene is used to produce a protein or part of a protein, to specify which protein or part of a protein is produced or what function it performs;

(25) Whereas, for the purposes of interpreting rights conferred by a patent, when sequences overlap only in parts, which are not essential to the invention, each sequence will be considered as an independent sequence in patent law terms;
chain termination reaction (Sanger 1977\textsuperscript{8}) and ABI Prism\textsuperscript{®} sequence analysis platform, both registered trademarks with Applera Corporation. A patent search in USPTO with Applied Biosystems as the assignee returned 79 hits related to automated sequencing and base calling software and instrumentation including robotic interface, sequencing reagents, reactions and dye chemistries. Applied Biosystems (ABI Prism) not only got its inventions protected but has also acquired licenses on several reagents that are used in sequencing.\textsuperscript{9} The emerging Pyrosequencing\textsuperscript{®} technology\textsuperscript{10}, relying on sequencing by synthesis, in contrast, is based on chain synthesis and the technology is owned by BioTage AB (Sweden). There are, however, many more patents and ownerships when it comes to sequencing methods and dye chemistries. Table 1 provides a glimpse of varied patents, both product and process related to sequencing methods, chemistries and reagents.

In so far as sequencing technologies are concerned, these may primarily be defined as stand alone tools and techniques developed for the researchers, the end-

<table>
<thead>
<tr>
<th>Table 1-Representative list of patents granted by USPTO on DNA sequencing related innovations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent</td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Sequencing dye chemistry based patents</strong></td>
</tr>
<tr>
<td>Fluorescent dye intermediates</td>
</tr>
<tr>
<td>Atropisomers of asymmetric xanthene fluorescent dyes and methods of DNA sequencing and fragment analysis</td>
</tr>
<tr>
<td>DNA sequencing using multiple fluorescent labels being distinguishable by their decay times</td>
</tr>
<tr>
<td><strong>DNA sequencing methodologies</strong></td>
</tr>
<tr>
<td>DNA sequencing method</td>
</tr>
<tr>
<td>DNA sequencing by Mass spectrometry via exonuclease degradation</td>
</tr>
<tr>
<td>DNA sequencing by parallel oligonucleotide extensions</td>
</tr>
<tr>
<td><strong>Reagents and equipment for sequencing reactions</strong></td>
</tr>
<tr>
<td>Polymer mixtures for high performance/high temperature separation in capillary electrophoresis, especially for long read DNA sequencing</td>
</tr>
<tr>
<td>Automatic processing system for use in solid phase biospecific binding and DNA sequencing techniques</td>
</tr>
</tbody>
</table>

(Source: Patent archive at www.uspto.gov)
users. The innovations are mainly aimed at reducing the cost of the tools and therefore directly impact on research and development constructively. The incentives to the innovators come forth in the form of royalty and licensing fee. Because these innovations are not of a derived nature, they do not pose any major patent thicket problem to the researchers.

**Sequence per se (Genomic, EST, GSS)**

The output of high throughput sequencing is mainly represented as Expressed Sequence Tags (ESTs), cDNA, DNA genomic survey sequences (GSS) and several other useful sequence categories such as Sequence Tagged Sites (STS), Single Nucleotide Polymorphism (SNPs), promoters and other regulatory factors and genes (Table 2).

*Sensu stricto*, DNA falls in the category of natural substances that already exist in nature and therefore should not be patented. However, the USPTO and EPO have ruled that ‘purifying’ or ‘isolating’ substances from their natural milieu requires manual intervention, which satisfies the novelty criterion making isolated DNA sequences a patentable subject matter. In theory, patent applications filed for DNA sequences may be drafted intelligently to support a broad range of claims. For instance, a disclosure on a partial gene fragment can potentially be linked to additional claims for the protection of the full-length gene, its complimentary nucleic acid sequences, and the translated protein. The supplementary claims may also include related genes that share a sequence identity of up to 90% and hybridize under specified hybridization conditions. This, therefore, implies that genes belonging to same or closely related families may be protected through patent claims issued on a single member. One such example relates to floral repressor gene *FLC*. The gene was isolated and characterized from model plant species, *Arabidopsis thaliana*, where its role as a floral repressor was established (Michaels and Amasino 1999).

<table>
<thead>
<tr>
<th>Patent</th>
<th>Patent number (Year)</th>
<th>Assignee</th>
<th>Inventor</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA sequencing by Mass Spectrometry via exonuclease degradation</td>
<td>EP689610 (2002)</td>
<td>Sequenom, Inc</td>
<td>Koester Hubert (US)</td>
</tr>
</tbody>
</table>

(Source: Espacenet® patent archive at http://www.european-patent-office.org)
authors claimed priority from US provisional patent application no. 60/121,572 filed on 25 February 1999 and US provisional patent application S. no. 60/123,455 filed on 9 March 1999. Subsequently, the inventors isolated the homologs from *Brassica rapa*, an economically important plant species related to *Arabidopsis thaliana* and obtained a patent. Examination of the patent document reveals that the claims drafted are extremely broad so as to cover not only the *FLC* orthologs but also paralogs from *Arabidopsis thaliana* and *Brassica rapa*.

The claims have a potential of blocking other parties from protecting the related genes of *FLC* family if they happen to share more than 80% similarity at amino acid level. This claim is critical because crop *Brassicas* are represented as diverse group of plant species. From an evolutionary point of view, the genomes of crop *Brassicas* are known to contain multiple copies of the same genes due to their polyploid origin.

In general, claims have to be supported sufficiently by ‘written description’ to satisfy specific, credible and substantial utility criteria and revised Interim Utility Guidelines Training materials, the US Patent and Trademark Office.

Contentious issues in patentability of ESTs came forth when the applications lacking any apparent ‘utility’ were filed. The ‘utility’ claims were often limited to the use of ESTs for isolation of full-length genes. Often the usage of the phrase ‘comprising of’ and ‘consisting of’, in the articulation of claims defined ‘open-endedness’ or ‘close-endedness’ of the scope of the patent, respectively. The phrases ‘comprising of’ afford a broader protection to even sequences non-recited in the patent document and may potentially result in infringement lawsuits. For example, claims articulated around ESTs by the first party using ‘comprising of/including’ tend to cover the yet-to-be isolated full-length ORF. This ORF when isolated later by second party would then infringe upon the claims of the first party. In contrast, ‘consisting of’ claims are narrower and more specific. Taking lessons from such cases, USPTO ruled that it will no longer grant protection to ESTs containing ‘comprising of’ claims and would support the claims that are narrower and specific.

As recently as 2005, the patents for ESTs were still being granted even when no apparent utility in terms of functional characterization of the full-length gene was demonstrated. This was a major concern for the companies engaged in microarray design and fabrication, which rely on the EST sequence information since that would entail obtaining licenses for every distinct EST represented on the array. The decision of the USPTO was affirmed by the Federal Circuit Court in *Dane K Fisher & Raghunath v Lalgudi* (2005); and this resulted in a paradigm shift in setting new standards for granting patents on ESTs. It was resolved that ESTs with no demonstrated ‘well-defined, real world benefit to the public’ can not be granted patent protection. This case has had the most significant impact on the use of ESTs with yet unproven function and utility in designing microarray, as firms engaged in microarray fabrication and design need not approach each and every researcher who has isolated an EST for license.

The third important issue involves those gene sequences that are anticipated to code for a protein whose function is inferred from the homology that it shares with another functionally characterized protein. Despite the resistance from researchers, the revised ‘Utility Examination Guidelines, 66 Fed Reg 1092 (5 January 2001)’ and the revised Interim Utility Guidelines Training materials, USPTO has indicated the following:

1. A protein whose function may be anticipated by virtue of its similarity to an already known protein, may be patented,
2. A sequence for a protein whose function is not known is unpatentable, and,
3. A sequence of protein which is known to bind a specific protein where the resulting dimer has no apparent function, is unpatentable.

However, these are only guidelines and cannot be construed as a statement of law. These are therefore not binding on the Court of Appeals for Federal Circuit (CAFC).

Another class of genomic sequences having wide-ranging applications concerns SNPs and other diagnostic markers. The availability of sequence data from various organisms has allowed researchers to examine the sequences threadbare and establish the link between mutation and genetic disorders, if any. These linked markers are extremely useful in predicting the predisposition of individual to a specific genetic disorder and in this respect have a greater technical effect as compared to ESTs. The development of a DNA sequence into a diagnostic
marker fulfills the criteria for IPR owing to its utility in diagnostics and thus can be granted patent. The standing examples relate to \textit{BRCA}1 and \textit{BRCA}2 mutations predisposing the carrier to breast cancers. Myriad Genetics (USA) holds the patent for use of the human \textit{BRCA} gene as diagnostic marker and enjoys sole monopoly on the use of \textit{BRCA} marker.\textsuperscript{19,20} In other words, it prohibits any other company to use \textit{BRCA}1 and \textit{BRCA}2 to predict predisposition to breast cancer. Other examples illustrate the patenting of SNPs where the specific utility aspect has been really vague and restricted to forensic identification purposes. In the current scenario, such inventions are unlikely to be patented, however, SNPs with a definite diagnostic utility are likely to be patented.

Finally, an aspect related to patenting of gene sequences and their derived products posing considerable ambiguity concerns splice variants wherein a single gene sequence encodes multiple messenger RNAs, each translating into a protein with potentially distinct functions.\textsuperscript{21,22} A case in point is the scenario with alternative splicing, wherein the original patent holder would have patented a gene sequence with a protein to the technical effect of having a specific function. However, another scientist may identify a splice variant originating from the same gene sequence but with a distinct function.

In the US, until now genomics related patents have been treated under traditional chemical patents law.\textsuperscript{23} Under this law, a patent on a novel chemical covers all the applications of the uses of that chemical whether discovered or not discovered, by the original patent holder. A discoverer of a new use may have rights to file a further patent, claiming use of the chemical for a particular purpose but would still require obtaining a license from the original patent holder before he commercializes his product. However, it is increasingly being realized that it is not appropriate to apply legal principles that work for chemistry to hold true even in genomics. The case in alternative splicing is not analogous to chemistry-based patent where the invention specifies a single chemical whose applications could be manifold. Hence, it has generally been observed that for the gene sequences giving rise to multiple splice forms, each isoform is a unique product and has to be claimed distinctly. For example, Savitzky \textit{et al} (2004) reported their invention in USPTO relating to alternative splice variants of CD40 wherein they claimed the novel nucleic acid sequences as well as the corresponding amino acid sequence.\textsuperscript{24} Earlier this year, Savitzky \textit{et al} (2006) filed yet another patent in USPTO, this time relating to six additional variants.\textsuperscript{25}

\textbf{Genomic Technologies}

\textbf{Management and Analysis of Sequence Data}

This part of the article focuses on an aspect of genomics, which aims at acquisition, processing, storage, distribution, analysis, and interpretation of high throughput sequence data all of which are described collectively as bioinformatics. As an integral component of genomics, bioinformatics integrates principles of mathematics and statistics to extract biological meaning from the wealth of data. In order to deal with the massive volumes of data, the application of computational methods become imperative.

Bioinformatics not only includes algorithms and software, but also data acquisition and storage devices. Maschio and Kowalaski (2001)\textsuperscript{26} have listed three basic types of bioinformatics innovations that may be considered for patent protection:

- Software algorithms, database architecture and hardware.
- Method of data acquisition from various sources (biological).
- The interpretation of data/information (i.e. product) that can be exploited.

These, of course, have to meet the criteria of patentability. One may argue that such hardware as microchip controlled devices (computers) have been in existence long before bioinformatics developed as a field. However, the novelty lies in the fact that the software now controls the hardware in a very specific and novel way. For example, a software algorithm may regulate the hardware that is involved in running a sequencing gel or a microarray-printing device. Coupled with this, the same hardware-software combination is involved in acquiring the raw data, converting into a readable and analysable format, and then stored on a physical device. It is this combination of hardware and software that may be considered for patent protection.

The following account describes the key components of bioinformatics:

- Sequence databases
- Sequence analysis tools

\textbf{Sequence Databases}

Databases, that store and organize sequence information electronically or non- electronically in a
systematic and easily accessible format have played a vital role in the genomics boom. The databases are not only the repository of curated information but also serve as start-point for several new endeavors. The databases in genomics came into existence with the establishment of the Los Alamos Sequence Database at the Los Alamos National Laboratory (NANL, USA) as a computerized repository of DNA sequence data. Gene-bank, currently managed by National Center for Biotechnology Information (NCBI) 27 evolved from the NANL database. Today, NCBI is among the largest repository of various genomics data that is freely accessible to the users. Besides, European Molecular Biology Laboratory (EMBL, Heidelberg, Germany), DNA Databank of Japan (DDBJ, Japan), Protein Data Bank (PDB, The State University of New Jersey, USA), Eukaryotic Promoter Database (EPD; Switzerland) represent other sequence databases in public domain. Apart from these, organism-specific, analysis specific or application specific, publicly available databases abound in plentiful. Some of these include TAIR (The Arabidopsis Information Resource), FlyBase (Drosophila resource), Gramene (Rice/Cereal database) (Table 3).

The databases in strictly private domain are fewer and exemplified by those maintained by Celera genomics, Incyte Corporation, Monsanto Corporation, Genome Therapeutics Corporation and Compugen Inc.

In the US and Canada, the databases may be patented only to the effect that the data itself qualifies for patent protection. Otherwise, databases per se may be protected through copyrights. However, genomic databases cannot be considered at par with any other database as they involve use of complex and high-end interrogation and interpretation algorithms and tools. Additionally, biological databases may require creating novel software-hardware architecture. With this background, the WIPO proposed a ‘Treaty on Intellectual Property in Respect of Database’ 28, which however did not culminate in formulation of a database legislation. Similar initiatives in the US have also not resulted in formulation of any guidelines. Notwithstanding, currently the European Directive provides guidelines to afford protection to databases through a dual mechanism of sui generis protection and copyrights. 29

There is a clear dichotomy in so far as access to the databases is concerned. In the preceding sections and Table 3, it is evident that the databases in public domain far out-number those represented by the private. This primarily reflects the open and collaborative spirit of research essential for its progress. However, from an industries’ perspective, seeking to capitalize on its intellectual property invested in generation and management of databases, several mechanisms exist that provide alternatives to non-existing patent protection. End user agreements, privacy and confidentiality clauses, licensing agreements, encryption of databases and physical restriction afford statutory protection. These forms of protection have been popularly adopted since both provider and the end user protect intellectual property. In contrast, patent clauses impose a mandatory public disclosure.

Sequence Analysis Tools

The raw sequence data has to undergo a preliminary analysis before it is organized as curated and/or annotated format in databases. This entails using algorithms and software for generation of primary information, which in itself is of no apparent utility. Further analysis of the primary data employs more sophisticated algorithms and software to provide structural and functional identities. The direct outcome of secondary analysis represents the true product of bioinformatics, an important component of functional genomics. The secondary information is of paramount importance since the knowledge on gene structure and function may be translated into industrial applications.

In general, legislations and guidelines regarding the patentability of software-related inventions vary significantly in different jurisdictions. In some countries, software is not considered as an invention unless it is linked to a ‘technical character’. 30 In other jurisdictions however, software as such generally constitutes patentable subject matter.

The US Federal Circuit rules that mere abstractions, ideas, mathematical principles and theories do not form a patentable subject matter. Nevertheless, the algorithms and software with a demonstrable practical utility are patentable as per USPTO guidelines. These applications include utility of algorithms and software as data mining, interrogation and interpretation tools.
Historically, software patents were not granted in the US. However, the decision of US Federal Circuit Court of Appeals on 'State Street Bank & Trust Co v Signature Financial Group Inc' led to amendments in the existing laws vis-à-vis patenting of software. Briefly, in the ‘State Street Case’ the USFCA held that algorithms that provide ‘useful, concrete and tangible result’ could be patentable.31

### Table 3—Representative databases and analysis tools

<table>
<thead>
<tr>
<th>S No</th>
<th>Database</th>
<th>Type</th>
<th>Protection status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nucleic Acid Sequence Databases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Genbank (<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>), NIH, Bethesda, USA</td>
<td>Nucleotide, protein, genome, structure, SNP, homogene, unigene, CDD, 3D domains, unISTS, GEO profiles, GEO datasets, cancer chromosomes, Genstat, dbSTS</td>
<td>Free access to database and analysis tools</td>
</tr>
<tr>
<td>2</td>
<td>EMBL Sequence Database (<a href="http://www.ebi.ac.uk">www.ebi.ac.uk</a>), EBI, Cambridge, UK</td>
<td>Microarray, nucleotide, protein, proteomic and structure</td>
<td>Free access to database and analysis tools</td>
</tr>
<tr>
<td>3</td>
<td>DDBJ (<a href="http://www.ddbj.nig.ac.jp/">http://www.ddbj.nig.ac.jp/</a>), Mishima Japan</td>
<td>DNA, Protein</td>
<td>Free access to database and analysis tools</td>
</tr>
<tr>
<td>4</td>
<td>EPD, <a href="http://www.epd.isb-sib.ch">http://www.epd.isb-sib.ch</a></td>
<td>Eukaryotic promoter database</td>
<td>Free access to the collection of promoter sequences and analysis tools</td>
</tr>
<tr>
<td>5</td>
<td>HUMHOT <a href="http://www.jncasr.ac.in/humhot">http://www.jncasr.ac.in/humhot</a></td>
<td>Database of Human meiotic recombination hotspots</td>
<td>Free access</td>
</tr>
<tr>
<td>6</td>
<td>MiRBase, <a href="http://microrna.sanger.ac.uk/">http://microrna.sanger.ac.uk/</a></td>
<td>MicroRNA sequence and target database</td>
<td>Free access</td>
</tr>
<tr>
<td>7</td>
<td>MODOMICS, <a href="http://genesilico.pl/modomics/">http://genesilico.pl/modomics/</a></td>
<td>RNA modification pathways</td>
<td>Free access</td>
</tr>
<tr>
<td><strong>Protein 3-D structure related databases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>PDB (<a href="http://www.rcsb.org/pdb">www.rcsb.org/pdb</a>)</td>
<td>Protein data bank</td>
<td>Free Access</td>
</tr>
<tr>
<td>2</td>
<td>SWISS-MODEL Repository</td>
<td>Automatically generated protein models</td>
<td>Free access</td>
</tr>
<tr>
<td>3</td>
<td>ModBase</td>
<td>Database of comparative protein structure models</td>
<td>Free access</td>
</tr>
<tr>
<td><strong>Protein sequence databases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Swiss-Prot, <a href="http://www.expasy.org/sprot/">www.expasy.org/sprot/</a></td>
<td>Protein sequences</td>
<td>Free access</td>
</tr>
<tr>
<td>2</td>
<td>Kabat, <a href="mailto:gjohnson@kabatdatabase.com">gjohnson@kabatdatabase.com</a></td>
<td>Proteins of immunological interest</td>
<td>One time licensing fee of US $2250.00</td>
</tr>
<tr>
<td>3</td>
<td>PRINTS, bioinf.man.ac.uk/dbbrowser/PRINTS/</td>
<td>Protein Motif Fingerprint</td>
<td>Free access</td>
</tr>
<tr>
<td>4</td>
<td>PANDIT, <a href="http://www.ebi.ac.uk/goldman-srv/pandit/">http://www.ebi.ac.uk/goldman-srv/pandit/</a></td>
<td>Evolution-centric database of protein and associated nucleotide domains</td>
<td>Free access</td>
</tr>
<tr>
<td>5</td>
<td>MIPS, mips.dsf.de</td>
<td>Protein sequence database</td>
<td>Free access</td>
</tr>
<tr>
<td><strong>Expression related databases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>dbEST; <a href="http://www.ncbi.nlm.nih.gov/dbest">www.ncbi.nlm.nih.gov/dbest</a></td>
<td>Database of Expressed Sequence Tags</td>
<td>Free access to public and data submitted by Genesifter customers</td>
</tr>
<tr>
<td>2</td>
<td>GeneSifter (<a href="http://www.genesifter.net">www.genesifter.net</a>)</td>
<td>Microarray data</td>
<td>Free access to raw data</td>
</tr>
<tr>
<td>3</td>
<td>AtGenExpress (<a href="http://www.tair.org">www.tair.org</a>)</td>
<td>Gene expression data from microarray experiments in <em>Arabidopsis</em></td>
<td>Free access to analysis tools</td>
</tr>
<tr>
<td>4</td>
<td>Genevestigator (<a href="https://www.genevestigator.ethz.ch/at/">https://www.genevestigator.ethz.ch/at/</a>)</td>
<td>Gene expression data from <em>Arabidopsis</em></td>
<td>Free access to analysis tools</td>
</tr>
<tr>
<td>5</td>
<td>The Institute of Genomic Research (<a href="http://www.tigr.org">www.tigr.org</a>)</td>
<td>Databases and analysis tools</td>
<td>Free access</td>
</tr>
</tbody>
</table>
The European Patent Office also follows largely similar laws with respect to patentability of algorithms and software.

A characteristic feature of genomics especially bioinformatics is its integrative and collaborative nature which is pivotal for its promotion and evolution. For this reason, the science of bioinformatics has by-and-large adopted an ‘Open-Science and Open-Source’ philosophy where unrestricted access and sharing of knowledge prevails. For example, the Gene-bank database at NCBI was established as a portal where researchers could deposit their sequence data for the entire scientific community to benefit from. This necessitated the development of database interrogation software such as Basic Local Alignment Search Tool\(^{32}\) permitting researchers to conduct homology searches and sequence mining. BLAST and many other related tools, in the spirit of open-science was placed in the public domain. The assignment of databases in the public domain led to three desirable outcomes. First, the Genbank database grew exponentially and became a single focal point for worldwide scientific community to contribute, share and mutually benefit. Second, it spurred the development and refinement of several sequence analysis and interrogation tools such as BLAST\(^{31}\) and CLUSTAL.\(^{33}\) Third, as the database grew, the power of finding the best matching hit to the query sequence increased manifold. It is to this effect that one witness that large number of databases and analysis tools can be freely accessed as depicted in Table 3 and 4.

### Genomic Technologies

#### Microarrays for Genome Analyses

Microarray technology is based on the principle of micro-fluidics, which relates to a host of processes and devices involving fluid kinetics at nano-or picolitre scale. Schena \textit{et al}\(^{34}\) define microarray as ‘an ordered array of nucleic acids, proteins and small molecules that enables parallel analysis of complex biochemical samples’. In essence, microarrays represent possibly the very first applications of nanotechnology.

Chip technology represents one of the most powerful genomics tools since it leads to the generation of genome-wide high-throughput data in a parallel manner on gene expression profiling, DNA polymorphisms, methylation status and alternative splicing states.\(^{35,36,37,38,39}\) Besides, it has created novel approaches for genetic mapping through detection of eQTLs.\(^{40,41}\) The chip technology is at the apex of genomics and integrates knowledge from all the other co-ordinates of genomics. The power of microarray technology is strengthened immensely through inputs from both public and privately owned databases (Table 3). Besides, a whole lot of software tools are required right from the array synthesis to data acquisition, processing, analysis and storage. Some of the commonly used microarray software is provided in Table 5.

Table 4-Representative list of Sequence analysis software

<table>
<thead>
<tr>
<th>S No</th>
<th>Software</th>
<th>Free access/Proprietary</th>
<th>Vendor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DNASIS™</td>
<td>Proprietary</td>
<td>Hitachi/Miraibio</td>
</tr>
<tr>
<td>2</td>
<td>GCK™ and GeneInspector</td>
<td>Proprietary</td>
<td>Textco</td>
</tr>
<tr>
<td>3</td>
<td>Sequencher™</td>
<td>Proprietary</td>
<td>GeneCodes</td>
</tr>
<tr>
<td>4</td>
<td>VectorNTI™</td>
<td>Some modules free for academics/commercial</td>
<td>Accelrys</td>
</tr>
<tr>
<td>5</td>
<td>Accelrys GCG (formerly GCG Wisconsin)</td>
<td>Proprietary</td>
<td>EBI/EMBL</td>
</tr>
<tr>
<td>6</td>
<td>EMBOSS Artemis</td>
<td>Free</td>
<td>EBI/EMBL</td>
</tr>
<tr>
<td>7</td>
<td>Free</td>
<td>EMBL</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>MEGA3.1©</td>
<td>Free for research and education</td>
<td><a href="http://www.megasoftware.net/">http://www.megasoftware.net/</a></td>
</tr>
<tr>
<td>9</td>
<td>ClustalX/ClustalW</td>
<td>Free</td>
<td>EBI/EMBL/several others</td>
</tr>
<tr>
<td>10</td>
<td>ElDorado/Gene2Promoter</td>
<td>Proprietary/free for academic use</td>
<td>Genomatix</td>
</tr>
<tr>
<td>11</td>
<td>PAUP* 4.0</td>
<td>Proprietary</td>
<td>Laboratory of Molecular Systematics</td>
</tr>
<tr>
<td>12</td>
<td>PHYLIP</td>
<td>Free access</td>
<td>University of Washington, Department of Genetics</td>
</tr>
</tbody>
</table>

Miniaturization of the biochip through microfabrication relies extensively on technologies from semi-conductor industry. The latest chip designs strive to accommodate and ever growing number of sequence arrays at highest density. For example Affymetrix high-density probe arrays can offer ollig densities as high as \(10^6\) per cm\(^2\).

The microarray as a research tool is being commercialized by several vendors such as Affymetrix Corp (Santa Clara, CA, USA), Incyte
Pharmaceuticals Inc (Palo Alto, CA, USA), Synteni Inc (Palo Alto, CA, USA), Hyseq Inc (Sunnyvale, CA, USA) and Agilent Tech (USA). The undisputed leader in this field is Affymetrix Corp, which earns close to US$ 350 million from annual sales of microarray and related reagents. As per the current estimates, the total market for microarray and related articles is set to cross over US$ 1 billion in the next couple of years. Keeping up with the heterogeneous demands regarding specifications of chip composition, customized arrays are now being developed and commercialized. The entire gamut of activities associated with microarray design, experimentation and analysis requires pooling of varied scientific skills is fraught with patent thickets. The proprietary inputs encompassing microarray technology are illustrated through authors’ experiences primarily from over 305 patents issued to Affymetrix for microarray technology. The patents issued to Affymetrix maybe categorized into five distinct types discussed below. Besides these, there are 55 odd patents that may be classified as of general nature.

**Microarray Synthesis and Manufacture**

Affymetrix has nearly 75 patents vis-à-vis arrays per se and various technologies for array manufacturing and synthesis. Of these, nearly 55 relate to prominent technologies and methods for oligonucleotide array synthesis using photolithography; combinatorial strategies for polymer synthesis, methods of printing oligonucleotide arrays and methods and use of photoactivable silane compounds. Affymetrix IP portfolio also includes patents regarding methods of arraying probes at high densities. Other methods for array manufacturing relying on micro-spotting devices and inkjet printing however belong to other companies.

**Microarray Platform**

Various patents cover the hardware component of Affymetrix chip arrays. Most of this hardware is automated to ensure precision and high-throughput data analysis. The microfluidic station for washing and signal detection come with associated proprietary software. Around 20 patents relate to miscellaneous devices of fluidics station and uses thereof including apparatus and design of chip packaging device, cartridge washing system and methods, hybridization device and method. Around 12 patents relate to integrated nucleic acid device, miniaturized genetic analysis systems and methods, among several others.

Other patents relate to scanning and imaging system with associated software. For example the GeneChip HP Scanner system by Hewlett Packard is protected by 32 patents. The US Pat No 5,578,832 relates to method and apparatus for imaging a sample on device, US Pat No 5,631,734 for detection of fluorescently labeled materials and US Pat No 6,171,793 for method of scanning gene probe array to produce data having a dynamic range exceeding that of scanner.

**Microarray Reagents and Quality Control**

The Affymetrix assays and reagents are covered as close to 30 patents related to chemical reagents and dye chemistries. These include methodologies and reagents for labeling of target through in-vitro transcription resulting in amplification of cRNA that is used as labeled target. Two representative patents include those reciting the use of functionalized silicon compounds and methods for their synthesis, nucleic acid labeling compounds. Four patents in the IP portfolio of Affymetrix relate to the use of quality controls as methods for testing oligonucleotide arrays.

**Microarray Analysis Software**

Nearly 45 patents belonging to Affymetrix relate to software used for hardware operation as well as data analysis tools. For example US Pat No 5,733,729
relates to computer-aided base calling for arrays of nucleic acid probes on chips. US Pat No 6,090,555 relates to scanning image alignment methods, US Pat No 6,185,561 concerns methods and apparatus for providing an expression mining database, US Pat No 6,223,127 relates to polymorphism detection utilizing clustering analysis, US Pat No 6,611,767 teaches scanned image alignment systems and methods and US Pat No 6,965,704 instructs system method and computer software for grid alignment of multiple scanned images.

**Microarray Applications**

The patents issued by Affymetrix teach the experimental applications of array in DNA sequencing, DNA fingerprinting, chromosomal mapping and specific interaction screening. Around 25 patents describe the applications of chips in genotyping. Another 15 patents relate to the use of chips in expression profiling.

**Patent Thickets in Microarray Technology**

The very first patents covering microarrays were issued in early 90s and were obtained by Hyseq, Affymetrix, Oxford Gene Technologies (Oxford, UK) and Stanford University (CA, USA). As the technology developed, the giants in the field sought to define their respective IP space resulting in a series of infringement lawsuits.

**Hyseq v Affymetrix**

Hyseq pioneered SBH (Sequencing by Hybridization) wherein oligonucleotide probes with lengths ranging from 11-20 nucleotides are arrayed in an overlapping manner, each array varying slightly from the other. These are hybridized to target nucleic acid whose sequence is to be interrogated. Based on the hybridization pattern, which permits discrimination of perfect matches and mismatches, the sequence of the target maybe determined enabling sequencing in a high throughput format. The patents obtained by Hyseq were directed to the method of sequencing DNA using microarrays. Affymetrix on the other hand pioneered the microarray synthesis and manufacture and obtained a series of key patents. A series of several other patents were later granted to Affymetrix. One important patent application was issued to protect the use of a set of mismatch and perfect match probes in order to interrogate the sequence of unknown fragments apart from many other claims.

A series of litigations ensued in March 1997 wherein Hyseq filed lawsuit against Affymetrix for infringement of its patents US Pat No 5,202,231, US Pat No 5,525,464 and US Pat No 5,695,940. Not only did Affymetrix defend itself from the infringement charges, but also claimed that Hyseq had infringed on its patent (US Pat No 5,744,305). The US District Court of the Northern District of California rejected all the claims made by Hyseq regarding infringement of its patents by Affymetrix. As a consequence, all the pending lawsuits were dissolved as an out-of-court settlement and the patents owned by both the companies were held valid. Further, the two companies not only cross-licensed their respective technologies but also went ahead to forge a tie as N-mer Inc for promoting the technology of sequencing by hybridization.

There have been similar infringement lawsuits between Oxford gene technologies versus Affymetrix and Incyte versus Affymetrix indicating that the microarray technology presents a complex patent landscape. Coupled to this is an observation that the magnitude of licensing issues in respect of chip technology can be enormous. An average chip may represent close to 40,000 sequences, which in most cases are genes. Considering a scenario wherein each and every spot is patented, the microarray manufacturers would be forced to obtain licenses for each spot which may render the enterprise economically non-feasible. In fact, each domain of microarray technology is covered with multiple derived patents and the only plausible solutions can come from collaborations and cross-licensing. Simultaneously, government intervention maybe required to frame appropriate guidelines that impact on microarray technology.

**Indian Position in the Field of Genomics and IP**

The Indian Genomics, especially in plant genomics is poised to expand through Indian Initiative for Rice Genome Sequencing (IIRGS) as part of the International Rice Genome Sequencing Project (IRGSP). Their activity forms a nucleation point in India for several other national and/or international programmes in the area of genomics. However, the R & D is still limited to short term projects and the emphasis is largely on publications and not IP protection. This is apparent as the Indian Patent Office received only 88 patent applications from 1995 to 2003, with the word ‘DNA’ in its title. The
intellectual property scenario in terms of genomics may thus not be described as coming of age in the very near future. However, the awareness of patents is percolating at a slow yet steady pace among the scientific community and in India as a whole. This possibly is an outcome of the recent amendments in the Indian Patent Act which has created curiosity and interest in every Indian Scientist, as reflected in the growing patent portfolio of CSIR.

The Indian Patent Office recognizes that in the field of biotechnology an invention may be related to living entity of natural origin, such as animal, plant, human beings including parts thereof, living entity of artificial origin, such as microorganism, vaccines, transgenic animals and plants etc., biological materials such as DNA, Plasmids, genes, vector, tissues, cells, replicons etc., process relating to living entities, process relating to biological material, methods of treatment of human or animal body, biological process or essentially biological process.

- The biological material such as recombinant DNA, Plasmids and processes of manufacturing thereof are patentable provided they are produced by substantive human intervention.
- Gene sequences, DNA sequences without having disclosed their functions are not patentable for lack of inventive step and industrial application.
- Any biological material and method of making the same which is capable of causing serious prejudice to human, animal or plant lives or health or to the environment including the use of those would be contrary to public order and morality are not patentable such as terminator gene technology.
- In case of use of biological material in the invention disclosed in the patent application the source or geographical origin of such material is required to be mentioned in the specification. In case of use of new biological materials in the invention disclosed in the patent application, such materials are required to be deposited in any of the International Depositary Authorities (IDA) recognized under the BUDAPEST Treaty on or before filing of the application in order to supplement the description for sufficiency of disclosure of the invention and reference of such deposit to be made in the patent specification.

With the above guidelines in place and the Indian Patent Act becoming more liberal in terms of accepting biotechnological inventions, India has started attracting foreign companies to start building their patent portfolios in India as well. The Annual report of Indian Patent Office shows a significant jump in the biotechnological inventions filed for patent protection in the last few years.

Despite the encouraging trend, it remains to be seen how genomics related innovations will be received in Indian market. The evolving Indian IP scenario coupled with strict enforcement will be critical in promoting the interests of various stakeholders, while promoting research and development.

References
1. The invention per se should constitute an appropriate subject matter. The USPTO (US codes, Title 35, Patents Section 101) rules ‘any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof’ forms an appropriate subject matter for IP protection. This can be interpreted to ‘cover anything under the sun made by man’ (Diamond v Chakrabarty, 1980).
2. The novelty clause emphasizes that the invention should not have been disclosed in the ‘prior art’. Publication of results by way of public disclosure at a meeting or conference, article appearing in journal, agreeing to license or sale one year prior to filing a patent application leads to non-fulfillment of novelty clause.
3. The non-obviousness criterion instructs that the invention should not be obvious to the people ‘skilled in the art’. This criterion aims to define invention by quantifying what is already known as prior art and what is being patented. In order to fulfill this criterion, the new development/invention must not have been obvious to a person ordinarily skilled in the art.
4. The utility criterion is linked to its usefulness with respect to its industrial application. In the revised ‘Utility Examination Guidelines, 66 Fed Reg 1092 (5 January 2001) and the revised Interim Utility Guidelines Training materials, the US Patent and Trademark Office has also indicated additional criteria describing the utility aspect by stating that ‘A sequence which may be linked to ‘a substantial and specific utility that is credible’ is patentable’. This doctrine, therefore strives to create a balance between the initial and subsequent innovations. The additional guidelines have also been devised to discourage frivolous claims in respect of utility for DNA sequences, which have not yet been characterized.
6. Eukaryotes are a group of living beings that have a nuclear membrane encompassing the genetic material, and comprise of fungi, plants and animals. This is in contrast to the prokaryotes, which lack well-organized nucleus, and are exemplified by bacteria and viruses. Eukaryotic genomes that have been sequenced or are in progress include human, mouse, frog, yeast, rice, poplar, Arabidopsis, papaya and soybean. Acidobacterium, Agrobacterium,
Anabaena and Bordetella are among the prokaryotes that are being sequenced, http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html.


For example, ABI has acquired the license to use AmpliTaq, the most thermostable DNA polymerase for driving the chain termination reaction whose original patentee is Roche Molecular Systems and F Hoffman-La Roche Ltd (Abramson R D, Gelfland D H, Greenfield I L, Mutated thermostable nucleic acid polymerase enzyme from Thermus species Z05. US Pat No 5,455,170, (to Hoffman-La Roche, Inc USA), 3 October 1995.


Genentech Inc v Chiron Corp 112F 3d 495, 501 (Fed Cir 1997).


Dane K Fisher & Raghunath v Lalgudi No. 04-1465, Slip Op at 11 (Fed Cir 7 September 2005).


Shattuck-Eidens D M, Simard J, Durocher F, Emi M, Nakamura Y, Linked breast and ovarian cancer susceptibility gene, US Pat No 5,693,473 (to Myriad Genetics, Inc USA; Centre de Recherche du Chul, USA; and Cancer Institute, Japan), 2 December 1997; Shattuck-Eidens D M, Simard J, Durocher F, Emi M, Nakamura Y Linked breast and ovarian cancer susceptibility gene US Pat No 5,709,999 (to Myriad Genetics Inc USA; Centre de Recherche du Chul USA; Cancer Institute, Japan), 20 January 1998.


Article 52 of the European Patent Convention (EPC); According to the Board of Appeals of the EPC, (T 1173/97 and T 935/97), the technical effect must go beyond the ‘normal’ physical interactions between program and computer. A ‘normal physical interaction’ includes display items on screen, store a particular pattern in a memory, activate a peripheral device, or at the very least, cause certain electrical currents to run over particular connections. A ‘technical effect’ can be reduced memory access time and therefore faster data retrieval and analysis (as required in databases and analysis tools), better control of an external device (such as a microarray printing robot or a colony picking tool). If such an effect can be found, the program is not excluded and hence a patentable invention.

State Street bank & Trust Co v Signature Financial Group Inc, 149 F 3d 1368 (Fed Cir 1998).


Finkelstein D, Ewing R, Gollub J, Sterky F, Cherry J M, Somerville S, Microarray data quality analysis: lessons from
44 www.affymetrix.com; 17 October 2006.
60 McGall et al US Pat No 6,596,856 McGall G, Barone A D, Nucleic acid labeling compounds, US Pat No 6,596,856, (to Affymetrix, Inc USA), 22 July 2003.


