A Critical analysis of the paper: "Intellectual Property Landscape of the Human Genome"

"Read not to contradict and confute; nor to believe and take for granted; nor to find talk and discourse; but to weigh and consider."
Francis Bacon, English essayist, "Of Studies," 1625.

By Sébastien Schauinger[1]

(e-mail: geneip@hgpr.org)

This paper sets out and discusses the result of a critical analysis of the Article **"Intellectual Property Landscape of the Human Genome**[2]**"** (hereinafter the "gene patent paper" or "GPP") and the conclusions we draw are indisputable: the GPP is flawed and does not provide any legitimate evidence that genes are highly patented. Indeed, there are numerous problems with methods and data interpretation in the GPP and this critical analysis will focus on those of greatest significance. Part 1 of this article provides a review of the gene patenting background and looks briefly at the definition of a gene patent. Part 2 discusses more specifically the reasons for the invalidity of the study design and the results of the GPP. Part 3 briefly reports the finding of the "true" human gene patent landscape as evaluated in the 2012 Human Genome Patent Report[3]. An appendix presents the different types of patent claims in the gene-related patents that can be found in the GPP.

Introduction

In October 2005, the journal *Science* published an article that allegedly purported to present an exhaustive analysis of human gene patenting[4]. Although almost everything is

---

[1] I am a European Patent Attorney. The opinions expressed in this article are my personal opinions and are provided to stimulate discussion.

[2] Jensen K, Murray F (2005) Intellectual property. Enhanced: Intellectual Property Landscape of the Human Genome Science 14 October 2005: 239-240.

[3] This report is accessible on the web site www.hgpr.org

[4] Fiona E. Murray, an associate professor at Sloan with a background in chemistry has been active in the intellectual property policy debate. She also submitted on January 19, 2010 a declaration in support of Plaintiffs in the Myriad Gene Patent Litigation.
She certified that her empirical research in this area demonstrates that, by using a precise algorithm based on bioinformatics methods, she identified *all* patents that claimed human nucleotide sequences (as listed using the REFSEQ language) in the claims and that her results revealed that **4382** of the **23,688** genes listed in the National Center for Biotechnology Information's gene database, or nearly 20% of human genes, are explicitly claimed (for some use or another) as United States intellectual property.
http://docs.justia.com/cases/federal/district-courts/new-york/nysdce/1:2009cv04515/345544/225/0.pdf

flawed and misleading in this now outdated paper, it received and still continues to receive considerable attention from scholars, "anti-gene patent" activists, and even from IP specialists. Because the GPP claims to fill the gap created by a lack of empirical data on the extent of gene patenting[5], and because such valuable information was long-awaited, the GPP succeeded in capturing a large audience.

The conclusion of the GPP that "20 % of genes are patented", not only permitted Association like the Council for Responsible Genetics (CRG) to urge policymakers to limit IP rights granted on human genes[6], but also permitted certain people to develop stupid rhetoric[7] built on an opportunistic and false interpretation of the GPP.

This is simply not acceptable for reasons of consistency and intellectual honesty.

I don't know if Jensen and Murray (hereinafter J&M) fabricated the statistics of their study with the intention of deceiving the public (we shouldn't forget that statistics are published to mean something and to be used for particular purposes). It is evident however, that J&M found what they expected to find and after carefully analysing their study, the only reasonable conclusion is that their intentions were clearly to cast a negative shadow over gene patenting[8].

I was also surprised that, given the outrageous degree of consideration J&M's Policy Forum published in Science received, nobody tried to highlight the weaknesses of

---

[5] According to J&M, most previous analyses have relied on anecdotal evidence (S. M. Thomas, M. M. Hopkins, M. Brady, Nat. Biotechnol. 20, 1185 (2002), R. Eisenberg, C. R. Biol. 326, 1115 (2003), M. Stott, J. Valentine, Nat. Biotechnol. 21, 729 (2003), J. P. Walsh, A. Arora, W. M. Cohen, Science 299) and empirical analyses have been hindered by:

(i) limited (and poorly defined) coverage of DNA sequence patents (M. Stott, J. Valentine, Nat. Biotechnol. 21, 729 (2003) and G. Xu, A. Webster, E. Doran, World Patent Inform. 24, 95 (2002));

(ii) difficulty separating patents that claim gene sequences per se from those merely disclosing DNA sequences (G. Dufresne, M. Duval, Nat. Biotechnol. 22, 231 (2004) Georgetown University Kennedy Institute of Ethics, DNA patent database and National Genome Information Center, Patome database).

[6] The Genomic and Research Accessibility Act proposed by Congressmen Becerra and Weldon in 2007. http://www.geneticsandsociety.org/article.php?id=3119)

[7] Michael Crichton in *The New York Times*: "the human genome exists in every one of us, and is therefore our shared heritage and an undoubted fact of nature. Nevertheless 20 percent of the genome is now privately owned. The gene for diabetes is owned, and its owner has something to say about any research you do, and what it will cost you". http://www.nytimes.com/2006/03/19/opinion/19crichton.html?pagewanted=all.

A number of misunderstandings about patents and gene patents appear to exist for many people (the average "man on the street" or even a PhD audience) through miscomprehensions on IP rights. For example, a patent on a gene sequence does not equate to ownership of that sequence; it does not confer ownership of the physical material as it exists in the body. The internet is full of inane comments such as those uttered by Drew Halley and David Koepsell, the author of "*Who Owns You*".

[8] J&M used the artificial and meaningless concept of "hot spots of heavy patent activity"(i.e "*BMP7 and CDKN2A were the most highly patented genes in the genome, their sequences were each claimed in 20 patents" and PSEN2, 8 assignees for 9 patents are the signs searched by the authors to expose the expansive gene patenting such that no single entity has enough liberty to pursue its program of research and development*").

methodology and the irrelevancy of the results of the GPP. I recall that after reading the GPP in 2005, I could not stop thinking that everything was wrong with the GPP.

However, J&M have made it difficult to evaluate or understand the real meaning of their questionable statistics by failing to document the data they obtained. Contrary to the usual practice of making all data freely available in order to facilitate study replication by others[9], J&M declined my request to share the full data of their study. Thus, I had no ammunition at my disposal to "pull the plug on" that damned "20 % of genes are patented". My life continued, and in 2008 and very recently, Prof Holman's articles[10] provided me with the missing material to prove the existence of numerous untruths in the GPP. This paper finally is the result of a thorough and objective analysis of the GPP. Frankly, I believe that the toughness of my words are on a level with the objective in view.

Before identifying the GPP's main flaws, I provide a brief review of the gene patenting background and an explanation of what is meant by a "gene patent".

PART1

1.1 Background to gene patenting

Since the early 1990s, (i.e at the beginning of the gene-hunting era), when the human genome project (HGP) was initiated[11], thousands of new genes have been isolated and sequenced.

It was then accepted by the whole research community that the availability of this new genetic information could enhance the discovery and development of new drugs[12]. But,

---

[9] J&M simply did not fulfil their responsibilities of authorship; all data necessary to understand and assess the conclusions of the manuscript must be available to any reader (general information for authors). http://www.sciencemag.org/site/feature/contribinfo/index.xhtml

[10] Christopher M. Holman, Trends in Human Gene Patent Litigation, Science 10 October 2008 Vol. 322 no. 5899 pp. 198-199 and Debunking the myth that whole-genome sequencing infringes thousands of gene patents, Nature Biotechnology,Vol.30 ,240–24, 2012.

[11] Begun formally in 1990, the U.S. Human Genome Project was a 13-year effort coordinated by the U.S. Department of Energy and the National Institutes of Health. Project goals : *identify* all the approximately 20,000-25,000 genes in human DNA, *determine* the sequences of the 3 billion chemical base pairs that make up human DNA, *store* this information in databases, *improve* tools for data analysis, *transfer* related technologies to the private sector, and *address* the ethical, legal, and social issues (ELSI) that may arise from the project.
http://www.ornl.gov/sci/techresources/Human_Genome/project/journals/journals.shtml

[12] *"The concept of the "druggable genome" has been used to denote a list of genes that can serve as suitable targets for developing therapeutic drugs. Drug discovery comprises a number of stages that lead from a biological hypothesis to an approved drug. Target identification is typically the starting point of the modern drug discovery process. A protein is highlighted through experimental techniques if it is shown to cause disease, either due to a change in its behaviour brought about by e.g. a mutation. Today, about 5,000 to 10,000 potential drug targets in humans for small-molecule drugs and antibodies were predicted to exist*". John P. Overington, Bissan Al-Lazikani and Andrew L. Hopkins, *Nature Reviews Drug Discovery* **5**, 993-996 (December 2006).

due to a massive patenting trend in the competitive biotechnology area and given the potential risk of restricted access to genomic research tools (such as genes), some pharmaceutical companies were "practically forced" to file patent applications on genes, at least to preserve their freedom to operate. For example, Merck & Co has taken the position to file patent applications on genes to ensure continued R&D without requiring rights from third parties[13]. Merck & Co also organized the Merck Gene Index Project (MGIP) to make cDNA sequences from various organs and tissues available to all scientists for gene identification and indexing. Merck & Co has sponsored a group at Washington University to sequence the human genome and make the information freely available.

In the same manner, GlaxoSmithKline believes that the practice of issuing patents in the field of genetics could promote research in this area[14]. SmithKline Beecham has formed a consortium with Human Genome Sciences to map, sequence, and patent as much of the human genome as possible[15]. Companies such as Human Genome Sciences (HGS), Millenium Pharmaceuticals and Incyte Pharmaceuticals sold genetic information to drug companies such as SmithKline Beecham, Roche Holding AG, Eli Lilly & Co, and Bayer AG[16].

Also, "to patent or not to patent" has been the main dilemma facing the different University-affiliated genome Centers and Research Institutes[17] involved with the HGP.

As a result, when the private U.S. firm Celera Genomics and the publicly funded "International Human Genome Sequencing Consortium" published[18] the first draft sequence of the human genome, the number of DNA-based patents was already high[19]. This was

---

[13] According to Mr. Tribble, Merck's IP position regarding the genomic research tools was to negotiate a non-exclusive license for rights on reasonable terms. Gene Patents, A Pharmaceutical Perspective Jack L. Tribble, Cambridge Quarterly of Healthcare Ethics (1998), 7:429-432 Cambridge University Press.

[14] GLOBAL PUBLIC POLICY ISSUES GlaxoSmithKline's position.
http://www.gsk.com/policies/GSK-on-proposals-for-a-disclosure-in-patent-applications.pdf

[15] Patenting gene sequences Caplan and Merz 312 (7036): 926 BMJ (1996)

[16] http://www.wsws.org/articles/2000/apr2000/gene-a10.shtml

[17] The main HGP genome centers payed particular attention in the dissemination of scientific knowledge and generally followed a non-patenting policy. The Whitehead Institute of Biomedical Research (WIBR) and The Baylor College of Medicine-Human Genome Sequencing Center (BCM-HGSC) used to patent genes, However, The Wellcome Trust Sanger Institute (WTSI), Washington University Medical School Genome Sequencing Center (WUGSC), University of Washington Genome Center (UWGC), Stanford Human Genome Center (SHGC), DOE Joint Genome Institute (JGI) did not allow gene patenting.

[18] Venter, J. Craig, et al. 2001. "The Sequence of the Human Genome," *Science* **291,** 1304-1351 and The International Human Genome Mapping Consortium *Nature* 409**,** 934-941 (15 February 2001).

[19] See the DNA Patent Database (DPD), an online database of DNA patents compiled and administered by the Kennedy Institute of Ethics at Georgetown University which focuses on patents disclosing DNA or related terms associated with DNA or nucleotide sequences whatever the species or the organism. The DPD contains a collection of DNA-based patents and patent applications issued by the United States Patent and Trademark Office (USPTO). Patents included in the DPD are identified by a search algorithm that captures documents with a nucleic-acid specific term in its claims section. Terms specific to nucleic acids include "DNA," "RNA," "nucleotide," "polynucleotide," etc. The DPD contained more than 22,000 patents in 2001. As of February 2, 2010, the DPD contained 52,716 issued patents.
http://dnapatents.georgetown.edu/aboutdpd.htm

mainly due to the explosion of the trend to patent nucleotide sequences on a mass scale, with or without a clear knowledge of the function and the future utility of these nucleotide sequences[20].

1.2 Human gene number

The first draft sequences revealed that the human genome contains somewhere between 30,000 and 35,000 protein-encoding genes, though some scientists, and most science writers, thought there were about 100,000 genes. Current estimates[21] give the number of protein-coding genes in the human genome as 21,724 known genes. These proteins may be classified according to their biological processes[22] or more conveniently according to their real life applications such as 1) therapeutically active compounds as such (ie.g tissue plasminogen activator, epoietin, growth hormone), 2) therapeutic targets (the so-called "research tools" i.e. the GPCRs, the kinases), and 3) DNA sequences useful for diagnostic or prognostic tests (i.e Thiopurine Methyltransferase, BRCA1 where polymorphs/mutants of DNA sequences are associated with diseases), for the most.

Also, the term "gene" may be used in a broader sense to encompass not only protein-encoding genetic sequences, but other structural or functional regions of the genome as well. According to *Wikipedia*[23], the term "gene" includes regulatory and other functional regions and also transcribed regions that are not subsequently translated into protein (transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), microRNAs or dsRNAs).

The Office of Technology Assessment (OTA)[24] has defined the terms "*DNA* sequences" and "*gene* sequences" in the following manner. "*DNA sequences and genome sequences* (the broadest terms used by OTA), *refer to an ex vivo length of an ordered sequence of nucleotides, generically; whether an in vivo function or association is known, putatively suggested, or unknown is not implied by the use of these terms*". In contrast, "*gene sequences is used to refer to those DNA fragments for which a specific biological product or in vivo function has been demonstrated and associated with the ordered sequence of nucleotide base pairs*".

---

[20] "Some genomics companies hurried straight to the patent office to stake a claim on any and all newly discovered genes without any understanding of their probable function. It has been alleged that some biotech companies routinely downloaded the freely available data produced by the Human Genome Project every night and filed a patent claim immediately, without making an effort to learn anything about the genes they were claiming intellectual property rights to". The biotech investor's bible, George Wolf, June 2001.

[21] Ensembl release 61 - Feb 2011 (Known + novel protein-encoding genes: 20,935+615), Pseudogenes: 13,483, RNA genes: 8,383, Immunoglobulin/T-cell receptor gene segments: 553, Gene exons: 603,260, Gene transcripts: 167,074. http://www.ensembl.org/Homo_sapiens/Info/Index.

[22] Proteins may be grouped functionally as involved in metabolism, ion transport, protein-phosphorylation, regulation of transcription, protein biosynthesis, cell structure etc…

[23] A gene is "*a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions, and or other functional sequence regions*" cited from Pearson H (2006). "Genetics: what is a gene?". Nature 441 (7092): 398–401. http://en.wikipedia.org/wiki/Gene.

[24] http://en.wikipedia.org/wiki/Office_of_Technology_Assessment

For the purpose of the GPP, J&M have chosen to consider the term "gene" as a short-hand for the set of proteins-coding genes[25]. The definition retained by J&M does not include i) regulatory regions, ii) functional RNA that is not subsequently translated into protein.

Having set the limitation of the term "gene" as used in the GPP, we can now move on to the definition of what constitutes a "gene patent" or a "patented gene" within the GPP. As seen, because there are a finite number of protein-coding genes in the genome, it would be very interesting if not important to evaluate the number of patents that specifically claim those sequences[26].

1.3 The extent of the gene patent landscape

In 2000, a special report[27] on the ethics of genetics identified the top 10 patentees of human genes worldwide. Leading the list was Genset, a French company, which has applied for patent applications on more than 36,000 human gene sequences. When looking at these numbers, the reader has to be very careful. All these numbers of gene sequences tell us very little about how many genes are patented or how many patent applications were filed or how many patents were actually granted. Moreover, what are exactly "gene sequences"? What is the correlation between the top gene patentees, patent applications filed and patents granted? This example shows how the artificial deployment of meaningless numbers can generate the illusion of extensive patenting of genes.

Since the beginning of the 2000s, several lists[28] purported to rank the best gene patenting organizations. Not surprisingly, all these "number lists" are different because no accurate definition is provided in order to adequately frame the gene patent landscape. Generally, these kinds of articles are always reluctant to provide any explanations about how the numbers were retrieved. At best, these kinds of "gene number lists" provide great

---

[25] This category (i.e the gene patents claiming human protein-encoding nucleotide sequences) is the subject of our analysis of the patent landscape of the human genome" (see page 239, first paragraph of the GPP). However, J&M's methodology was irrelevantly designed to retrieve fragments of genes such as immunogenic epitope or even oligonucleotides of less than 30 nucleotides (see page 10 for analysis).

[26] It should be noted that even if J&M defined the term "gene" as "human protein-encoding nucleotide sequences", they only searched patent disclosing nucleotide sequences in claims and not "claims on genes defined through amino acid sequences"(see page 13 for analysis).

[27] The top 10 patenters on the human body: Genset has applied for patents on 36,083 gene sequences, Ribozyme (US) claims patent applications on 15,863 sequences, Genetics Institute (a subsidiary of American Home Products) has applied for patents on 9,876 sequences, Genzyme (US) has filed applications on 8,546,Hyseq (US) claims 6,147 sequences,Human Genome Sciences (US) has filed for patents on 3,964,US Department of Health 2,991 sequences,Affymetrix (US) 2,079 sequences,Genentech (US) 1,995 sequences,Incyte (US) claims 1,755; The top 10 human gene patenters claim 70.4% of the total of 126,672 human gene sequences between them. http://www.guardian.co.uk/science/2000/nov/15/genetics8.

[28] The Great Gene Grab, September 2000 By Antonio Regalado. http://www.technologyreview.in/biomedicine/12184. The number of gene patents granted each year has shot up almost 14-fold since 1990. Data include patents on human, animal, plant and microbial gene sequences. Top 10 patent holders: Incyte Genomics 397, University of California 253, Glaxo SmithKline 248, U.S. Dept. of Health & Human Services 205, Novo Nordisk 196, Genentech 165, Isis Pharmaceuticals 146, Chiron 135, American Home Products 130, Novartis 128.

headline material, but not much else. But statistics are a favorite evidence of many writers and speakers who use various numbers as such in support of their ideas and their conclusions. Incyte and Human Genome Sciences, two of the most active gene patentees used to manipulate the numbers of gene patents, when they were racing with each other to be the number one in the patenting of genes.

Incyte described itself as the leading genomic information company, allegedly having over 375 U.S. patents on human genomic structures (essentially three times as many as any other organization) and claiming Incyte has another 6,500 patent applications pending[29]. But, have those numbers anything to do with the reality?

Similarly, William Haseltine, CEO of Human Genome Sciences (HGS) spoke on a *Motley Fool Radio Show*[30]. Here is a little extract of what he had to say about his business: "*We are definitely the king of the gene patent. We are in first for several reasons. We were the first genomic company. We discovered virtually all the human genes by mid-1995. Human genes are something you can only discover first. There is not an endless supply. What people are doing now, the Celeras of the world and other companies and the government, are rediscovering what we already found. We have very, very good evidence that as early as 1995 we had isolated and characterized in our laboratories and in our [Human Genome Sciences databases] 95% of all human genes.We have isolated, characterized, and sequenced more than 120,000 different human genes. So the number is certainly greater than 120,000. Those people, who estimate lower, simply don't know. We have isolated and characterized them so we're in a very good position to know that there are more than 120,000 of them. Now the next thing we've done, and this distinguishes us from all other genome companies and projects, is we actually go out and make the product of the gene and determine what it does in living biological systems. We have now filed patents that describe over 7,500 human genes and their medical uses*."

So it can be said that an important aspect of Incyte's or HGS's strategy was the acquisition of proprietary rights to new genes and proteins. Although that would seem to put Incyte or HGS in a powerful position (they could have claims on up to 30% of total human genes), the IP gene patent position of Incyte, more than 10 years after those demonstrations of impressive enthusiasm about the gene patenting business is anything but excessive[31].

---

[29] The Wall Street Journal, March 16, 2000, Michael Waldholz.

[30] March 6, 2000,http://linkage.rockefeller.edu/wli/reading/hgs.html

[31] *Supra* note 10. Prof. Holman reported in 2012 that "*as of April 2011, only 37 of the 398 Incyte patents flagged as gene patents in J&M article were still in force, the others had all expired owing to Incyte's failure to pay the necessary maintenance fees*". In fact, already by 2001, Incyte had exited the following activities: microarray-related products and services, genomic screening, products and services, public domain clone products and related services, contract sequencing services, transgenics products and services and single nucleotide polymorphism ("SNP") discovery services (see the 2003 Annual Report). On February 2, 2004, Incyte announced the cessation of development of the information products. Today, Incyte is focused into Small-Molecule Drug Discovery and Development.

1.4 What does and does not constitute a gene patent

When you analyze data, you usually assume that you know what the data really represent. Or do you? For the purpose of their article, J&M choose to define a gene patent as "a patent that claims a human protein-encoding nucleotide sequence"[32]. In other words, the goal of J&M was to assess how many genes were patented per se.

Numerous patents have been granted in relation to biotechnological inventions which claim genetic material in some or other manner. However, to determine what does and does not constitute a gene patent is certainly not an easy task, and certainly demands to understand how to "play" with the patent claims.

There are various possible definitions of what constitutes a "gene patent". Much academic research uses the term "gene patent" to refer to any patent dealing indistinctly with the full or a fragment of the gene sequence, only disclosed in the description of the patent or specifically claimed by the patent.

Yet in 1998, Heller and Eisenberg[33], two ardent opponents of gene patents, argued that patenting human genes may deter innovation. They used the case of "adrenergic receptors" as an illustration of their theory and found over 100 patents that might require, according to them, but wrongly, a license to do research in this area. Responding to Heller and Eisenberg's appealing but unfounded theory, Seide and MacLeod[34], two patent specialists, did a search on "adrenergic receptors" and, indeed, found 135 patent documents. They commented as follows: "*Although brief, our review indicated that the majority (of patents) would not be infringed by such an assay, thus obviating the need for a license. For example, NIH has patented a method of treating schizophrenia by administering an alpha2 - adrenergic receptor antagonist and a D2 dopamine receptor antagonist to a patient. Although this patent has the term "adrenergic receptor" in the claim, the assay to detect a ligand to such a receptor would not be covered, and no license need be obtained. Thus, in order to determine whether a license to any or all of the issued patents would be required in order to search for suitable ligands, one would be able to eliminate many on a first reading. In reality, licences to only certain technology would be required and these can be negotiated with the owners*". Finally, "the numbers of owners and potentially blocking[35] patents has been overstated "said Seide and MacLeod.

---

[32] *Supra* note 25

[33] Michael A. Heller, Rebecca S. Eisenberg (1998) Can Patents Deter Innovation? The Anticommons in Biomedical Research; *Science* 1 May 1998 Vol. 280. no. 5364, pp. 698 – 701

[34] Rochelle K. Seide, Ph.D. and Janet M. MacLeod, Ph.D; http://www.sciencemag.org/site/feature/data/980465/seide.xhtml

[35] The concept of a blocking patent stands for the principle that a patent, for example covering a product is infringed no matter how the product is produced or used. A blocking patent can also be defined with regards to its ability not to be easily designed around. A blocking patent is a patent which cannot be worked without falling within the scope of protection of a blocking patent. In fact, in the adrenergic receptor case, there are less than 10 blocking patents claiming different genes coding adrenergic receptors per se. Many of the other patents that have been counted only claim some narrowly defined method involving the use of the adrenergic receptors and do not restrict the use of the said receptors as such. Just

Similarly, J&M tried to abusively justify their position on the basis of a biased work. Indeed, J&M defined the term "gene patent" as “any patent disclosing and claiming a human gene sequence or some fraction[36] thereof”. Thanks to Christopher Holman’s disclosure[37], it is now crystal clear that J&M’s dataset of gene patents comprises entire or partial sequences, cDNA molecules produced in a laboratory corresponding in sequence to human mRNA molecules, but also non-naturally occurring nucleotide material with mutation, deletion, addition (fusion genes), antibody or antisense sequences which have nothing to do with natural protein-encoding genes. J&M’s dataset also comprises all other categories of patents (method of use, process) that cannot be considered as “gene patents[38]” according to the ultimate goal of J&M’s study which is “explicitly designed to be selective only for those patents that explicitly claim a sequence using the SEQ ID language with a sequence listing“. Therefore, J&M’s methodology failed to correctly distinguish a patent claim that only discloses a gene in it from a patent claim that actually claims a gene. There is no other way to conclude that by intentionally exaggerating the scope of their study, J&M create a false impression that “20% of human genes are explicitly claimed as U.S. IP”.

The next part reviews the key problems with the GPP. It is particularly argued that J&M’ study is filled with methodological and analysis problems that lead to the unjustified conclusion of 20% of human genes being patented.

We begin by examining critically the study design used by J&M to achieve their goal and we follow by reporting all major issues found in the GPP.

PART2

2.1 The study design

Each study design is the product of choices: the choice of definition, the choice of one measurement over another measurement and the choice of sample to emphasize some aspect of a problem. As this critical analysis will show, the choices made by J&M fatally undermine the conclusion of the GPP.

---

because a patent refers to such a gene or receptor in the claims does not necessarily mean that the gene or the receptor encoded by the said gene is claimed as an invention.

[36] http://www.sciencemag.org/content/310/5746/239/suppl/DC1. In supporting Online Material, J&M comment: “we included only those (patented sequences) that were at least 150 nucleotides in length on the basis that this is the average length of one human exon and yet still small enough to capture EST sequences.

[37] *Supra* notes 10 & 31

[38] See page 22 for all existing gene-related patents.

The research question addressed by J&M was discussed at the beginning of the GPP. J&M justified their study design on the basis of the "*lack of empirical data on the extent of gene patenting*". Considering that J&M clearly defined what they specifically wanted to learn or understand by doing their study (the patent landscape of the human genome), they nevertheless failed to implement the right method to collect the data (the patents) that may answer their research question (how many human protein-encoding nucleotide sequences are patented). Indeed, the automated method used by J&M was not a realistic or a reliable method for retrieving specifically patents that claim protein-encoding genes per se, because there is no direct relationship between searching for and finding a nucleotide sequence defined by the term "SEQ ID NO" in a claim and asserting that the claim actually protects the said nucleotide sequence. Thus, a perceived discrepancy exists between what was actually searched for and what should have been kept in J&M's database in relation to the defined research question. By relying only on automated matching, J&M could not possibly know what and why possibly important patents they searched for are missing or inadequately identified. Automated data collection reduces researcher time and effort as compared to human observation, but in return, further validation of data collected should be flawless. This is not the case for the GPP, because evaluating accurately the true gene patent landscape would have required taking into account other considerations J&M simply did not address and would have required an expenditure of research time that J&M simply did not undertake.

Errors, flaws and shortcomings occurring in the GPP are addressed one by one in the following paragraphs.

2.2 Misleading as regards the definition of what was searched

A fundamental problem with the GPP is its inability to clearly keep the same definition of what is a gene patent and what is not and why. The goal of the GPP was to evaluate the number of genes that are patented. J&M chose to define gene patents as "those claiming human protein-encoding nucleotide sequences". However J&M's dataset includes a lot of false positives resulting from an absence of manual checks.

The discrepancy existing between the goal of the search (counting specifically U.S patents claiming protein-encoding genes) and what was actually retrieved[39] completely negates the relevancy of the GPP. Also, in their "Supporting online material" (see in particular page 4 first paragraph), J&M assert "*that though relatively rare, patents contain claims directed towards sets of genes for use as diagnostics in DNA microarrays experiments and typically have a structure like "a plurality of probes selected from SEQ ID NO: 1-xx,xxx. These patents claim many gene sequences*". This is not correct. For example, claim 1 of U.S. Patent 6,500,938 does not claim any gene sequence but a composition comprising 1490 polynucleotide probes. J&M will be happy to learn that there is a world of difference

[39] J&M's dataset comprises patents claiming genetically modified sequences (mutant, fusion protein), microarray (composition of hundreds of probes), peptides, antibodies, antisense sequences and various methods using genes or proteins (see pages 22-30).

between claiming "*a plurality of probes selected from SEQ ID NO: 1-1490"* and "*a composition comprising a plurality of probes which are SEQ ID NOs: 1-1490*[40]".

As a consequence, the analysis presented by J&M is everything but "*extremely specific in its identification of patents that claim specific human protein-encoding nucleotide sequences*". If the 4 microarray patents belonging to Incyte were removed from J&M dataset, less than 2000 human genes would constitute J&M's human gene patent landscape. However J&M's dataset comprises many, if not a majority, of patents claiming other subject matter than protein-encoding nucleotide sequences per se.

Therefore, for at least the foregoing reasons, the GPP is fatally flawed.

2.3 Flaws in the method of data collection

J&M's method consists in three main steps[41]: Step 1 was used to download nucleotide sequences from the GenBank patent division. Step 2 was developed to determine which sequences correspond to human genes. Finally, Step 3 was implemented, according to J&M to identify patents "*that actually claim the nucleotide sequences explicitly in the claim language*". The procedure was based on software written by J&M for parsing the text of patent claims, and to identify nucleotide sequence defined by the term "SEQ ID NO".

However, J&M's method calls into question the use of algorithms for targeting the identification of patents that actually claim human nucleotide sequences. According to J&M, their bioinformatics method was supposed to capture all protein-encoding nucleotide sequences claimed in U.S patents, but excuse me, by simply relying on the term "SEQ ID NO" in the claims, an automated analysis is not likely to capture information about gene patenting per se. In fact, contrary to what was stated in the GPP, the algorithm used by J&M was far from being able to identify all patents claiming human genes. The mere presence of the "SEQ ID NO" in a claim does not necessarily mean that the protein-coding sequence is claimed as such[42].

Therefore, for at least the foregoing reasons, the GPP is fatally flawed.

---

[40] Claim 1 of Incyte patent US6500938 reads: "A combination comprising a plurality of polynucleotide probes, wherein said plurality of probes are SEQ ID NOs:1-1490". Also patents US6607879, US6492505 and US6183968 claim compositions comprising a plurality of polynucleotide probes comprising respectively a nucleotide sequence of SEQ ID NOs: 1-1508, 1-305 and 1-134. These compositions are particularly useful as hybridizable array elements in a microarray for monitoring the expression of a plurality of target polynucleotides. The microarray comprises a substrate and the hybridizable array elements.

[41] See in particular pages 1-3 of the GPP's supporting online material.

[42] Supra note 39.

2.4 Flaws in the relevancy of the data source

J&M used the patent division of the GenBank database as the main source of gene patents. However, one knows very little about the patent division of the GenBank database. No information is available on the way GenBank gets data from patent offices, the specific details on which sequences are transferred, the rationale behind that selection, and the schedule on which sequences are uploaded. Lack of transparency on this matter is unfortunate and is extremely harmful to the validity of J&M's methodology[43].

Considering the difference of content between GenBank and commercial databases such as USGENE (STN), Q-PAT (GenomeQuest) or SQIP, J&M's approach excluded the majority (>80%) of the nucleotide sequences that were important to consider for landscaping gene patents.

Also, missing patents were downloaded from the USPTO's Patent FullText and FullPage Image Databases database[44] using a patent search to retrieve only those patents containing both the phrases "SEQUENCE LISTING" and "SEQ ID" in the patent specification. J&M did not explain why this set of patents (11,081 patents corresponding to 109,766 nucleotide sequences) were missing from the GenBank download. The absence of justification is detrimental to the accuracy of the method.

Therefore, for at least the foregoing reasons, the GPP is fatally flawed.

2.5 Uncertainty in the method of matching human nucleotide sequences

Step 2 of the GPP represents a method to determine which of the sequences retrieved from GenBank and the USPTO databases correspond to human genes. BLAST alignments[45] were made to compare the entire database of disclosed sequences to the complete set of mRNA transcripts from the NCBI RefSeq database. Only those alignments with at least 150 nucleotides in length and an E-value of exactly zero were kept. Now, the lower the E-value, or the closer it is to zero, the more "significant" the match is. In other words, exact (100% identity) matched hits will probably have E-values of 0! It can however be questioned whether a BLAST alignment with an E-value of exactly zero is able to detect individual mutation (known to cause disease) or variation between people not necessarily associated with gene alteration (polymorphisms). It is known that the source of DNA from NCBI RefSeq database is composite, derived from multiple donors of diverse ethnic

---

[43] It is estimated that GenBank's gbpat division represents only 13.5% of the SQIP dataset. The database "SQIP" contained 49,684,446 unique nucleotide sequences and 3,287,071 unique protein sequences. Compare that to GenBank's gbpat division which contained 6,728,032 unique nucleotide sequences. http://sqipdb.wordpress.com/2010/02/01/all-the-patent-sequences-are-in-genbank-right-no/ SQIP is a commercial database of DNA and protein sequences derived from patents. http://sqipdb.com.

[44] See in particular pages 1-2 of the supporting online material.

[45] http://en.wikipedia.org/wiki/BLAST. In bioinformatics, **B**asic **L**ocal **A**lignment **S**earch **T**ool, or **BLAST**, is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

backgrounds[46]. Thus, it is probable that by considering only hits with the very high statistical significance (E=0) the price to pay could be very high: very close sequences could probably be missed by J&M's method.

The reliability of J&M's method is also questionable since chimeric genes[47] created through the fusion of two or more genes have been identified. This is not consistent with a very strict E-value cut-off!

Therefore, for at least the foregoing reasons, the GPP is fatally flawed.

2.6 Error in choosing to use only nucleotide sequences

J&M choose to limit their search "to patents using the canonical "SEQ ID NO" claim language without considering claims on genes defined through amino acid sequences. The reason given for this important defection clearly reflects a lack of knowledge of patents and particularly of gene-patents. J&M did not provide any explanation or evidence to substantiate this study design choice, except for a vague statement that "*most patent claiming amino acid sequences (or using claim structures such as "a nucleotide sequence encoding a polypeptide sequence in SEQ ID NO..") tend to disclose full-length cDNA sequences for these polypeptides, even if they are not claimed per se using the SEQ ID terminology*." It is hard to follow the logic of this reasoning. The fact is that J&M did not care about patents claiming polypeptide sequences and ignored a full category of patents claiming specifically human polypeptides. This choice is problematic, given that the paper is interested in the count of protein-encoding genes. Each gene can be claimed in different ways. In fact, it appears that a protein-encoding gene defined via its polypeptide sequence represents the preferred way of claiming genetic sequences in patents[48]. Therefore, a polypeptide sequence is probably the best indicator of the presence of a gene claimed in a patent. As patent drafting is very heterogeneous, it is logical to consider that a gene may be patented through either its polypeptide or its nucleotide sequence.

Moreover, considering the degeneracy of the genetic code[49] there is a virtually infinite number of nucleotide sequences that might code for a specific protein. In other

---

[46] http://en.wikipedia.org/wiki/Human_Genome_Project; Genome donors: HGP scientists used white blood cells from the blood of two male and two female donors (randomly selected from 20 of each) each donor yielding a separate DNA library. The most common form of variation is the single nucleotide polymorphism (SNP). Although two unrelated people share, on average, 99.9% sequence identity (i.e., one difference in a thousand base pairs), the average occurrence of an SNP in the general population is once every few hundred base pairs. As such, more than nine million unique SNPs have been cataloged in the public database, dbSNP (Crawford and Nickerson 2005) with many more expected to be found in large-scale resequencing efforts.

[47] See page 26, example 12

[48] See page 23-25, examples 2, 6-7, 9

[49] Given the four different nucleotide bases A, C, G and T, there are 64 possible codon combinations. Based on the mappings of codons to amino acids defined by the genetic code, 61 of these codons code for one of the 20 amino acids. The other 3 correspond to the special stop codons. The degeneracy of the genetic code allows for more than one codon to code for the same amino acid. A base substitution within a codon that does not result in a change in the encoded amino acid is known as a synonymous codon substitution. Non-synonymous codon substitutions correspond to base changes in a codon

words, a considerable variation in nucleotide sequences could translate into the same amino acid sequence. Thus, claiming a specific polypeptide sequence is a more convenient way to protect a gene product than a nucleotide sequence subject to more variations.

By excluding polypeptide sequences from their definition of genes, in all likelihood, J&M created a bias which in turn has certainly affected the results and invalidated the GPP's conclusions. Finding the right gene patent landscape was certainly not their number one concern[50].

Therefore, for at least the foregoing reasons, the GPP is fatally flawed.

2.7 Flaws in the sampling method and the problem of generalization

In order to determine the scope of their analysis, J&M conducted an analysis on a well-defined sample of patents known to be relevant in gene patenting. J&M chose to study the patent portfolio of the U.S Company, Human Genome Sciences (herein after HGS). According to J&M, the HGS patent sample is representative of "true" protein-coding gene patents. This assertion may or may not be true. In fact, the homogeneous picture of the HGS patent portfolio is somewhere miraculous for J&M's demonstration, because it creates the false impression of unity of the dataset. Indeed, one of the first questions that needs an answer is to know if the chosen sample[51] is representative of the larger population (the dataset). By sampling only HGS's patents, J&M excluded all other patentees[52]. Even if the study were otherwise flawless, valid generalization about the whole population of gene patents could not be drawn from the HGS sample. From a rational point of view, pretending that simply calling the dataset homogeneous will make it so, is a serious mistake. J&M did not establish any correlation existing between the whole dataset and the smaller subset of the dataset (the HGS sample). In order for this to work correctly, one thing has to be true: the sample must be similar to the target population (J&M's whole dataset) in all relevant aspects and this is far from being the case. The presence of heterogeneity of their dataset (distinct subpopulations of patents or patent claims) has simply not been analyzed and

---

that result in a change in the encoded amino acid. Missense and nonsense mutations produce non-synonymous codon substitutions.

[50] "*4,382 genes were claimed as intellectual property. This is probably an underestimate because the analysis didn't look at patents claiming rights to proteins and which might have omitted the gene sequence*", says Dr. Murray. The actual number "could be up to double" the 4,382, she says. http://online.wsj.com/article/SB112922829130567900.html

[51] Sampling is the act of selecting a suitable sample or a representative part of a population for the purpose of determining parameters or characteristics of the whole population. A sample is expected to mirror the population from which it comes, however, there is no guarantee that any sample will be precisely representative of the population from which it comes. Sampling error comprises differences between the sample and the population that are due solely to the particular units that happen to have been selected. For example, J&M chose as their sample HGS patents whose claims belong to the product-claim category and are constructed on the same scheme. It is very clear, even without any statistical proof, that this sample is a highly unrepresentative sample leading to invalid conclusions. This error can be detected very easily as shown by categorization of the patents of the J&M dataset. (see pages 22-30).

[52] Around 600 different patentees constitute J&M's dataset. Top 10 patentees counted for 1434 patents (33%), top 20 for 1893 patents (44%).

categorized so that all quantitative measures correspond to comparing apples with apples, not apples with oranges.

J&M have fallen into the trap of confirmation bias[53].

Therefore, for at least the foregoing reasons, the GPP is fatally flawed.

2.8 Flaws in quality control (Validation and Cleaning)

The assertion "*overall this analysis suggests that our method is extremely specific in its identification of patents that claim specific human protein-encoding nucleotide sequences*" is a pure myth because nowhere in the GPP can one find even the slightest evidence of the specific capture of protein-encoding gene patents per se.

J&M arbitrarily asserted that "*our analysis is explicitly designed to be selective*"[54]. Indeed, J&M reported that: "*as shown in the table, our manual reading of the HGS patent portfolio reveals claims directed towards 183 nucleotide sequences using the "SEQ. ID" language. Of these, our automated method finds 173, with no false-positives", but 10 false-negatives: 8 were coded nonhuman protein-encoding sequences, two false-negatives arose due to spelling and clerical errors*". Although this is true, this information resulted from a manipulation of the findings. Indeed, J&M (or whoever) read the patents for the purpose of properly identifying whether their automated method identified human rather than non-human genes in patents, but not to determine if they were actually "true"[55] protein-encoding gene patents.

Therefore, all this is very confusing, since J&M wanted us to believe that their automated method found 173 patents specifically claiming nucleotide sequences. However, a careful reading of the GPP only allows to conclude that the automated counting was based on first searching the term "SEQ ID" in the claims and on determining which of the identified nucleotide sequences actually corresponded to human genes (after step 3 of their method).

Again, we stress the fact that the method used by J&M was not capable of telling us whether a nucleotide sequence defined by its "SEQ ID" (human or not) is or is not specifically claimed!

---

[53] Confirmation bias is a tendency of people to favor information that confirms their beliefs or hypotheses.

[54] In this case, selectivity or specificity should be seen as the ability of the method used by J&M to correctly exclude patents that do not deal with human genes. Not surprisingly, as a bioinformatics method with E-value=0 was chosen to include with as much as certainty as possible a human gene, FP were absent. But, contrary to what was asserted by J&M, FP can easily be identified by reading the patents in J&M's dataset. J&M were playing loose with the facts and were also deliberately omitting everything that failed to support their findings that many genes are patented. Clearly, J&M were not interested in the truth and almost all readers were duped by the GPP.

[55] "True" protein-encoding gene patents should be understood as patents that specifically claim the nucleotides sequences using the SEQ. ID format, per se. In order to supposedly validate their method, J&M chose a test based on the ability to distinguish whether the retrieved patents deal with human or non-human gene patents.

Also, it should be noted that even a manual reading of the claims makes it impossible to conclude whether a given gene is human or not, unless the term "human" is explicitly present in the claim or at least in the specification.

Moreover, J&M reported that "*195 patents disclose but do not claim nucleotide sequences*"(those patents are not part of J&M's dataset because J&M only searched for patents having claims that recite nucleotide sequences under the form '"SEQ ID"). Very little was said about this full category of patents. Apparently, J&M did not know how to deal with the 195 other "true" protein-encoding gene patents! Or maybe J&M seemed to consider that these "polypeptide" patents are merely duplicates of the 183 nucleotides patents as members of patent families[56]?

In order to shed some light on the number of genes patented by HGS, we did our own analysis of the situation. By querying[57] the USPTO database, and after reading the claims of all retrieved patents, we identified a total of 321 human gene patent families. Of those patents, 179 patents contain at least one claim with nucleotide sequences defined by their "SEQ. ID NO:", 142 patents claiming nucleotide sequences or polypeptide sequences only via a polypeptide "SEQ. ID NO:". Our manual verification shows that of the 142 "polypeptide patents", at least 39 concern patents[58] that do not have a nucleotide patent counterpart. These patents are not found in J&M's dataset. Moreover, on checking the 173 HGS patents contained in J&M's dataset we identified a lot of redundancy. For example, the RBP5 gene has been claimed in at least 3 patents according to J&M's dataset[59]. The set of 3 patents

---

[56] A patent family refers to a group of equivalent patents or patent applications that cover a group of inventions related to the parent (i.e the first) patent (or application) in that family. Counting patents by family avoids the problem of double counting inventions that are patented in divisionals or continuations. A further complication arises when an original (or parent) patent application leads to additional divisional applications (or continuation and continuation-in-part (CIP) applications in the U.S). The idea that each invention is allowed to be patented once, is related to the idea that a patent can cover only one independent and distinct invention. Because the patent claims define the patent coverage or the scope of the invention, two separate patents can't literally claim the same invention." If it's the same invention, it can't be claimed twice." But, in case of gene patents, nucleotide sequences and polypeptide sequences are patentably distinct and may be two different claimed inventions, although they may be directed to the same gene (i.e invention). Counting patent families instead of all patents avoids the problem of duplication.

[57] Results of Search in US Patent Collection db for: (((AN/human AND genome) AND sciences) AND ISD/19900101-> 20050505): 486 patents. http://patft.uspto.gov/netahtml/PTO/search-adv.htm.

[58] 6,008,020; 5,994,302; 5,994,103; 5,986,069; 5,986,060; 5,981,230; 5,981,215; 5,968,797; 5,962,411; 5,958,729; 5,958,400; 5,945,321; 6,420,526; 6,632,920; 6,566,325; 6,147,050; 6,475,753; 6,495,520; 6,534,485; 6,576,445; 6,458,349; 6,534,630; 6,410,709; 6,844,170; 6,864,226; 6,537,966; 6,251,648; 6,495,128; 6,391,589; 6,319,700; 5,981,215; 6,489,138; 6,610,829; 6,783,971; 5,948,890; 6,759,512; 6,410,701; 6,881,823; 6,878,806; 6,566,498; 5,994,103; 6,667,390; 6,872,546; 6,537,539; 6,482,923; 6,486,301; 6,093,795; 6,139,832; 6,057,434; 6,046,031; 6,294,164; 6,623,941; 6,153,739; 6,569,992; 6,878,687; 6,605,592 ; 6,525,174; 6,365,369; 6,500,638; 6,806,351; 6,531,447; 6,342,581; 6,476,195; 6,590,075; 6,433,139; 6,534,631; 6,667,032; 6,448,230; 6,262,233; 6,372,473.

[59] The following genes were counted at least twice in J&M's dataset: RBP5 (5844081, 5977309, 6287812), ALKBH1 (5618717, 5747312), CST6 (5985601, 6617132), CST7 (5919658, 6066617), FGF17 (5728546, 6368822), FGF14 (5773252, 6013477), LGALS9 (6027916, 6468768), DNASE1L1 (5830744, 6569660), CTRC (5710035, 6107075), SLC18A1 (5859200, 5798223), TOB1 (6258777, 6013469), CCL24 (5880263, 6419917, 5866373), deoxycytidine kinase 2 (5914258, 6063376), MLH1 (6482606, 6620619, 6610477, TOP1MT (5723311, 6255077), geranylgeranyl pyrophosphate synthetase (5928924, 5786193), CMAS (6333182, 6783971), GPR37(6143519, 5750370), IMPA2 (5716806, 5955339), CTSK (5501969, 6383793, 6475766), PHYH2 (5945273, 6200796, 5635616), TIMP4 (6300310, 6448042), VEGFC (6608182, 5935820), CFLAR (6623938, 6680171), CASP7 (6087150, 6495519, 6538121), PON2 (5629193, 5792639, 6140093), CCL25 (5981231, 6503735), CORT

forming a group of patents linked to a single gene as for example RBP5, does not mean that this particular gene is patented 3 times. However, J&M would have us believe that this was the case.

Finally, our manual counting of HGS's patents revealed that at least **204** unique genes are patented. Of those, **127** patents were still in force as of July 2012.

Thus, J&M's methodology has produced a lot of false positives contrary of what was revealed by J&M! Considering all patents in J&M's dataset (product patents, use patents and process patents directed to genes, antibodies, antisense sequences and fusion proteins) as a homogenous set would be like "grouping dogs, chairs and fourth-degree equations under one category as 'quadrupeds' on the grounds that they all had four legs"[60].

J&M relied solely on automatic searches without manual examination of the patents and further cleaning of the data. Manual or automatic de-duplication is important because without this data cleaning step, certain patented genes may be over-represented in J&M's dataset. For example, the thirteen patents[61] assigned to the Institute of Virology of the Slovak Academy of Sciences concern only patents that claim, in thirteen different ways, the MN gene and its applications!

J&M encountered no problems in claiming that nearly 20% of genes are patented and believed that their analysis demonstrated that each of the 4270 patents can be adjudged as a gene patent. This is terribly wrong because biased data was captured and data was not analyzed properly. To succeed in their task and to be meaningful, J&M should have "qualified" the retrieved patents. The first qualification of a patent involves categorizing its claims so that all patents in the dataset correspond to comparing apples to apples, not apples to oranges. J&M did not check (or maybe they did, but did not care about it) the gene sequences claimed in the 4270 patents of their dataset. Remember, J&M only checked the claims of the 195 polypeptides and 173 nucleotides patents belonging to HGS. And, on that basis they tried to have us believe that the 173 very homogeneous patents of HGS can be generalized to the whole dataset of 4270 patents.

As explained above, calling the patents they identified "gene patents" would require putting more time and effort into quality control[62].

(6524826, 6232100), CCR5 (6511826, 6025154), CCL23 (6001606, 6451562), TNFSF13B (6689579, 6716576), NUDT5 (5695980, 6344547), CRADD (6130079, 6495322).

[60] This quote is from Robert Musil, Essays and Speeches, vol.8, 1921.

[61] US5955075, US5972353, US5981711, US5989838, US6004535, US6051226, US6069242, US6093548, US6204370, US6297041, US6770438, US6774117, US6027887. In fact, J&M did not consider only one patent from each patent family; on the contrary, from their methodology, J&M have arbitrarily and without any discernment taken all patents regarding the "same gene" belonging to the same patentee. It is therefore not clear whether the 4382 genes of J&M's dataset are unique or redundant counting.

[62] The following sentence "the top patent assignee is Incyte Pharmaceuticals/Incyte Genomics, whose IP rights cover 2000 human genes, mainly for use as probes on DNA microarrays" is symptomatic of J&M's lack of seriousness when manipulating patent data. It is standard practice for patent users to check the claims before asserting they claim

Last but not the least, J&M failed to report how many patents of their dataset were in fact still in force in May 2005. To adequately assess the extent of the gene patent landscape, it is an absolute necessity to take the patent renewal process[63] into account. J&M considered that each of the 4270 patents were in force. This is terribly wrong. Our estimate[64] is that more than 30% of the patents were abandoned.

All disclosed and undisclosed limitations or deficiencies of the GPP are summarized in Table 1. These limitations lead to either underestimation or overestimation of the number of gene patents. These limitations were also commented on in the foregoing paragraphs.
As can be seen, J&M selectively set forth findings that support their conclusion regarding gene patents and simply ignored the very limitations that did not support their conclusion that a large number of genes had been patented.

Underestimation	**Overestimation**
1/ The GPP concerns outdated information. However a patent landscape should be a dynamic work and should not be viewed as definitive. Not only the GPP represents a false snapshot of the situation in May 2005, but the patent applications still pending in 2005 and issued since the last seven years were ignored by J&M. MAJOR LIMITATION NOT DISCLOSED BY J&M	A/ Patents abandoned (not assessed by J&M) This limitation is CRITICAL in the appraisal of the number of patented genes. Unfortunately, this issue was overlooked. MAJOR LIMITATION NOT DISCLOSED BY J&M
2/ Protein-encoding genes defined by polypeptide sequences. (See comments pages 13-14). MAJOR LIMITATION DISCLOSED BUT NOT ASSESSED BY J&M	B/ Duplication of patents disclosing the same gene in patent families. This limitation was not assessed by J&M. MAJOR LIMITATION NOT DISCLOSED BY J&M
3/ Missing patents filed before 1990 (patents without nucleotide sequences in the computer readable format SEQ. ID). Many patents are absent from J&M's dataset. The first patent of the dataset is U.S 5,215,892 (priority date of 1989-12-25) granted in 1993/06/01. Our estimate is that about 170 gene patents were granted before 1993/06/01. MAJOR LIMITATION DISCLOSED BUT NOT ASSESSED BY J&M	C/ Nucleotide sequences only mentioned in claims but NOT CLAIMED. MAJOR LIMITATION NOT DISCLOSED BY J&M

something! To add insult to injury, J&M failed to report that a large majority of the Incyte patents had expired due to non-payment of maintenance fees!

[63] Maintenance fees or renewal fees are fees that are paid to maintain a granted patent in force. Maintenance fees on utility patents in the United States are due 3½, 7½ and 11½ years after grant of the patent.

[64] See the "Human gene patent report" disclosing that more than 1700 (out of around 5000) gene patents had expired due to nonpayment of maintenance fees.

4/ E-value of exactly zero (see comments pages 12-13) MAJOR LIMITATION DISCLOSED BUT NOT ASSESSED BY J&M	D/ Discrepancy between the study design (the objectives) and the results. Inaccurate data that do not fit the "gene patent" definition: fusion protein, mutant, antisense sequences, antibody, ESTs, peptides of less than 10 amino acids, method, use, composition (see comments pages 10-11). MAJOR LIMITATION NOT DISCLOSED BY J&M
5/ Spelling errors such as "SEQ D" or "SEQID" (admitted by J&M) This limitation roughly accounts for less than 20 patents. MINOR LIMITATION DISCLOSED BUT NOT ASSESSED BY J&M	E/ Misuse of the results: repeating that patent with SEQ ID in the claims is equivalent to patented genes. MAJOR LIMITATION NOT DISCLOSED BY J&M

Table 1: Disclosed and undisclosed limitations found in the GPP

In summary, J&M did not discuss or assess the major methodological limitations of their study, nor the effects of these limitations on the interpretation of their findings. Instead, J&M only cited minor limitations[65] and drew conclusions that cannot be supported by their findings. Indeed, J&M acknowledged that there were a number of "limitations" inherent in their research and specifically identified them as shortcomings to their paper. However, they offered no guidance as to the impact of these "limitations" on the reliability of their results. J&M's limitations are so huge that the various underestimations and overestimations may, paradoxically, cancel each other out to give a number not too far from the reality.

J&M took major shortcuts by simply capturing patents very arbitrarily. Drawing a conclusion based solely on blind counting is everything but serious. The GPP is based on analyzing wrong and incomplete data. Therefore, the GPP cannot be used to properly assess the IP landscape of the human genome.

Jensen & Murray have fallen into the same trap that they noted in other previous studies[66], by failing to distinguish between patents that merely recite genes in the claims and/or in the specification and patents that actually protect genes. The vast majority of the 4270 patents identified by J&M are not granted as "composition of matters" sequence claims. Most of the nucleotide or polypeptide sequences are not claimed directly but are included to support the claimed invention.

Therefore, for at least the foregoing reasons, the GPP is fatally flawed.

[65] See the paragraph "Caveats and notes," in the GPP's supporting online material.

[66] Supra note 5

2.9 Conclusion

A thorough reading of the GPP and of the auxiliary document (the supplementary online material), the re-analysis of the HGS patent portfolio and the checking of the whole dataset reveals that the GPP is a "bogus" study. We have shown in this critique that the result of the GPP falls apart under close inspection. Given its many flaws, the GPP patent counts cannot be considered as the foundation for a reliable paper. J&M's study is fatally flawed as it misrepresents data and offers unsupported conclusions.

Although a superficial reading of the GPP leaves the impression that 20% of genes are patented (remember J&M did not provide their patent dataset for public review and forced readers of the GPP to take their word for it, upon close analysis of the claims of the patents collected by J&M, it is apparent that the 20% statistic is nonsensical, and derived from misreadings, arbitrary choices and out-of-their-hat assumptions.

Some scientists and scholars have argued that "*having too many patents on a single gene stifles innovation because researchers hoping to work with that gene will think the risk of infringement is too high, or that licensing rights to the gene will be an uphill battle*"[67]. This is a myth. The fears raised by some regarding the negative impact of human gene patenting have been overstated. The 20% is jabberwocky! These digits are hardly more than decorative. But the solution to the problem of bad statistics is to become a better judge of numbers encountered in papers. Being critical means not being too credulous and not accepting any statistics at face value. When we fail to think critically, the statistics we hear or read might just as well be sensational. J&M's triumph was to create the concept of a "golden number" applying to gene patenting.

Determining what are the IP rights over a finite number of human genes is an important and challenging issue facing society, but since we are in a world dominated by sound-bites and scoop, reality and truth are frequent victims of easy misleading and sometimes completely inaccurate reporting. But what could be more serious than placing this kind of article in the hands of those who are not experts in the field of patents, but take advantage of opportunities to create misguided political decision? The GPP should not have been published because of its cumulative errors and misinterpretations.

Some consolation could be found in the fact that the GPP did not lead to bad policy choices being made, although the American Civil Liberties Union (ACLU)[68] and relatively unknown figures who invariably raise the same tired discourse about the Myriad case[69], are

---

[67] http://online.wsj.com/article/SB112922829130567900.html

[68] "The U.S. Patent and Trademark Office (PTO) has granted thousands of patents on human genes – in fact, about 20 percent of our genes are patented. A gene patent holder has the right to prevent anyone from studying, testing or even looking at a gene. As a result, scientific research and genetic testing has been delayed, limited or even shut down due to concerns about gene patents" (June 26, 2012).

http://www.aclu.org/free-speech-womens-rights/aclu-challenges-patents-breast-cancer-genes-0

[69] If you are interested in gene patents, the Myriad case is no longer a secret for you!

still blindly using the 20% gene patent quote to justify their absolute abomination of gene patents.

Hopefully, the open minded public has an interest in knowing that the GPP is fatally flawed. Now, it is done.

PART3

The Human Genome Patent Report

The "Human Gene Patent Report" (herein after HGPR) is a patent landscape on human protein-encoding genes.

A major goal of the HGPR is to highlight the gene patenting timeline from the first patent application filed[70] directed to a gene per se through to very recent filings. An objective of the HGPR is to illustrate the complexity of patent protection on human protein-encoding genes. A second objective of the report is to describe the search methodology used to perform in-depth analysis of patent literature related to human protein-encoding genes.

As it should be understood, the HGPR covers ONLY patents whose claims protect nucleotide or polypeptide sequences ("genomolecules") directed to protein-encoding genes (full or partial genes). This study does not deal with the large category of biotechnology patents, the so-called genetic patents that cover all patents whose claims include nucleotide sequences whatever their origins and uses.

The key findings of the HGPR are summarised below. Detailed findings are contained in the rest of the report downloadable from the website www.hgpr.org.

As of **August 2012**, the HGPR reports that:

-4977 gene patent families were granted by the USPTO,
-Of those patent families, 2964 are still in force,
-Genentech is the number one private patentee with 271 gene patent families in force,
-The top 10 patentees own 33% of all patents.

[70] The first gene-patent (US4447538) was filed in April 19, 1978 and granted in May 8, 1984, covered the human chorionic somatomammotropin.

Appendix 1

Appendix 1 identifies the different types of claims occurring in the "gene-related patents" that can be found in J&M's dataset (patents marked with an asterisk (*) are included in J&M's dataset).

A claim is a single sentence composed of three parts:

-the preamble identifies the category and the subject matter of an invention. The limiting constraints are found later in the body of the claim.

-the body lists the main elements of the invention, and

-the transitional phrase, including terms such as "comprising" or "consisting of", also important in assessing the scope of the claims as the phrase can be restrictive or permissive in nature.

An invention for which protection is sought, may also be set out after the wording "is characterized in that".

The first category of claim language that can be used in gene-based patents concerns patents covering a product or composition of matter. Claims can be directed to the genetic sequence *per se*, but also to a DNA probe capable of specifically hybridizing to and thereby recognizing a genetic sequence, or a specific mutation in the sequence. Claims can be directed to specific genetic constructs or expression systems, such as a recombinant vector, cell line, or host organism comprising the gene sequence. It is also possible to see the subject matter of a claim defined through the function performed by the claimed sequences such as the capacity of an antibody to bind an antigen.

The second category of patent covers method and use of genes or products of genes. For example, claims can encompass screening methods for identifying binding molecules of polynucleotides and polypeptides, methods and compositions for inhibiting or enhancing the production and function of the polypeptides.

Examples of these different types of claims include the following:

I. Product (composition of matter) claims

1. Isolated nucleotide or polypeptide sequences per se

Example 1 (claim 1 from US Patent No: 5,731,192*) reads:

-An isolated polynucleotide encoding a human alpha-6(IV) collagen

Example 2 (claim 1 from US Patent No: 5,744,349*) reads:

-A DNA sequence encoding the amino acid sequence of Myt1Hu, as set forth in SEQ ID NO:2

Example 3 (claim 1 from US Patent No: 6,166,191*) reads:

-An isolated and purified subgenomic polynucleotide comprising the nucleotide sequence shown in SEQ ID NO:1.

The most straightforward manner to claim a protein-coding gene is as follows: "an isolated DNA comprising the nucleotide sequence of SEQ ID NO:x". If SEQ ID NO:x is a fragment of the complete ORF of a given gene, this claim would be infringed even if the "accused" sequence contains element other than SEQ ID NO:x. However, according to the general practice of filing gene patents, a DNA molecule (an isolated nucleic acid) is very often indirectly claimed via its polypeptide sequence as shown in Example 2.

Example 4 (claim 1 from US Patent No: 5,559,023*) reads:

-An isolated DNA fragment comprising the full structure or a part of the DNA represented by SEQ ID NO:1, wherein said part of the DNA comprises a base sequence of at least 10 contiguous bases from the DNA represented by SEQ ID NO:1.

Example 5 (claims 1 & 2 from US Patent No: 6,379,950*) reads:

-An isolated nucleic acid molecule encoding a rev-caspase.

-The nucleic acid molecule of claim 1, wherein the rev-caspase is selected from the group consisting of rev-caspase-1, rev-caspase-2, rev-caspase-3, rev-caspase-4, rev-caspase-5, rev-caspase-6, rev-caspase-7, rev-caspase-8, rev-caspase-9, rev-caspase-10.

Each nucleotide sequence is presumed to represent an independent and distinct invention. However, up to ten independent distinct sequences can be examined in a single application without restriction requirement pursuant to 35 U.S.C 121 and 37 CFR 1.141.

Example 6 (claim 1 from US Patent No: 6,210,923) reads:

-Isolated and purified m-rigui2 protein, wherein said m-rigui2 protein has the amino acid sequence SEQ ID No. 3, and wherein said DNA is coded for by DNA selected from the group consisting of:

(a) isolated DNA which encodes an m-rigui2 protein;

(b) isolated DNA which hybridizes to isolated DNA of (a) above at conditions consisting of hybridization in 5 times SSC, 1% SDS at 65°C. followed by washing at 65 DEG C. with SSC ranging in concentration from 1 times to 0.1 times; containing 0.1% SDS, wherein said DNA encodes an m-rigui2 protein; and,

(c) isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to the degeneracy of the genetic code, and which encodes an m-rigui2 protein.

The “hybridization language” covers the particular sequence SEQ ID No. 3 originated from **mouse**, as well as any nucleotide sequence that hybridizes to that sequence under a given set of experimental conditions. It can reasonably be assumed that a human gene sequence will hybridize according to this claim 1. Therefore claim 1 of US Patent No 6,210,923 covers a human gene sequence of rigui2 protein.

Example 7 (claim 1 from US Patent No: 6,620,619*) reads:

> **- An isolated polypeptide comprising a first amino acid sequence 95% or more identical to a second amino acid sequence of amino acid residues 1 to 756 of SEQ ID NO:2 wherein said polypeptide has DNA mismatch repair activity.**

The percent identity language, which is also sometimes expressed as percent similarity language, covers not only the sequence of interest SEQ ID NO:2, but any sequence that is 95% identical to that sequence.

Example 8 (claim 1 from US Patent No: 5,935,814*) reads:

> **- An isolated polynucleotide that differs from the polynucleotide sequence of SEQ ID NO: 1 in that said isolated polynucleotide has up to a total of 3 nucleotide alterations per 100 nucleotides, wherein said nucleotide alterations are selected from the group consisting of: substitutions, and insertions.**

Example 9 (claim 1 from US Patent No: 6,335,435*) reads:

> **-A gene encoding:**
> **(a) a protein comprising an amino acid sequence of SEQ ID NO: 2; or**
> **(b) a protein having from one up to five amino acid deletions, substitutions or additions in the amino acid sequence of SEQ ID NO: 2, which has a helicase activity.**

A multitude of modifications can be made in a DNA molecule such as nucleotide deletion, addition, substitution without substantially altering the function of the encoded protein. In order to cover all possible embodiments, nucleotide or polypeptide sequence claims can be constructed to broaden their scope of protection.

Example 10 (claim 1 from US Patent No: 5,817,479*) reads:

> **- A purified polynucleotide having a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO: 4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, and SEQ ID NO:44.**

The subject invention provides unique polynucleotides (SEQ ID NOs 1-44) which have been identified as novel human kinases (kin). These partial cDNAs also called expressed sequence tags (ESTs) represent nucleotide sequence information from a short segment (usually between 150 and 500 base pairs (bp) of a randomly selected cDNA clone.

For example, SEQ ID NO:6 has 255 base pairs and refers to the clone 35652 HUVEC KEK5 (EPHB4 gene product). The entire gene has 4369 bp.

Example 11 (claim 1 from US Patent No: 5,298,407*) reads:

- Isolated DNA comprising a sequence encoding a polypeptide that is immunologically reactive with the monoclonal antibody produced by the hybridoma designated ATCC #HB 10319.

All nucleotide sequences that encode any amino acid sequence performing a specified function are covered by this king of claim.

Example 12 (claim 1 from US Patent No: 5,908,762*) reads:

- A chimeric DNA molecule comprising a DNA sequence that encodes a polypeptide having a sequence corresponding to SEQ ID NO:5, the sequence of said polypeptide selected from the group consisting of SEQ ID NO:5, muteins of SEQ ID NO:5, truncations of SEQ ID NO:5, and fusion proteins containing them, wherein said polypeptide specifically binds to double stranded DNA having the sequence of SEQ ID NO:3.

Example 13 (claim 1 from US Patent No: 5,843,888*) reads:

- A non-naturally occurring mutant human hemoglobin wherein the valine residue at position 96 of the alpha chain (SEQ ID NO: 1) is replaced by a tryptophan residue.

Examples 12 and 13 can be considered as very limited embodiments of the invention involving sequences SEQ ID NO:5 and SEQ ID NO:1 respectively.

Examples 1-13 recite claims that cover any and all of the sequence uses and cover the various constructs used for such uses.

The product *per se* claim is known to confer absolute product protection[71] because it carries no limitation, either as to the process by which it is made or as to the end-use to which it is employed.

In some case when a product patent does not exist, in the case of a research tool protein such as a GPCR, claims directed to a screening method may be considered as broad as a product claim because such claims cover specifically the main use of the product. Some claims are drafted in a manner that attempts to encompass any variant of a gene[72]. In some other cases, patents may claim more specific sequences[73].

A gene (a nucleotide sequence) can produce more than one protein product through alternative splicing or post-translational modification. A gene can produce more or less protein in different cells at various times in response to developmental or environmental signals, and many proteins can express disparate functions in various biological contexts.

2. Other products

Example 14 (claim 1 from US Patent No: 6,100,062) reads:

- An expression system comprising a polynucleotide capable of producing a HSCLOCK peptide comprising the amino acid sequence as set forth in SEQ ID NO:2 when said expression system is present in a compatible host cell.

Example 15 (claim 1 from US Patent No: 6,136,604*) reads:

- An antisense compound 8 to 30 nucleobases in length targeted to a start codon, nucleobases 1513-1884 of a 3'-untranslated region, or nucleobases 69-1414 of a coding region of human methionine aminopeptidase 2 (SEQ ID NO:3), wherein said antisense compound specifically hybridizes with and inhibits the expression of human methionine aminopeptidase 2.

[71] However see Judgment C-428/08, Monsanto Technology LLC. Since the Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the legal protection of biotechnological inventions (OJ 1998 L 213, p. 13) makes the patentability of a DNA sequence subject to indication of the function it performs, it must be regarded as not according any protection to a patented DNA sequence which is not able to perform the specific function for which it was patented. Therefore, because a DNA sequence such as that at issue in the main proceedings was not able to perform its function when incorporated in a dead material such as soy meal, claim 6 of the EP0546090 (B2), "a DNA sequence encoding a Class II EPSPS enzyme selected from the group consisting of SEQ ID NO:3 and SEQ ID NO:5" does in this particular case not confer an absolute protection.

[72] Examples 1, 4-11

[73] Examples 2, 3 and 12

J&M's algorithm identified more than one hundred patents belonging to Isis. It is well known that Isis was involved in the development of RNA-based technologies and successfully developed a drug discovery platform based on their antisense technology. By checking the claims of the Isis patents, it appears that almost all these patents found in the J&M dataset are directed to antisense sequences of less than 30 nucleotides. The sequences defined by the "SEQ.ID NO:" is in fact the gene targeted by the antisense sequence, the claimed subject matter of each patent.

Example 16 (claim 1 from US Patent No: 6,548,064*) reads:

> **- An isolated peptide consisting of from 8 to 12 amino acids, wherein said amino acids correspond to contiguous amino acids of an SSX molecule or NY-ESO-1, wherein said isolated peptide has a threonine residue or an alanine residue at both its second and terminal positions, and binds to an HLA molecule to form a complex.**

Example 17 (claim 1 from US Patent No: 6,534,268*) reads:

> **- An isolated nucleic acid corresponding to SEQ ID No: 1 that encodes a human bone morphogenetic protein-7 promoter region.**

Examples 16 and 17 are parts of J&M's dataset but according to J&M's criteria should not have been kept as gene patents.

Example 18 (claim 1 from US Patent No: 5,932,216*) reads:

> **- An antibody that specifically binds to a Bone Morphogenetic Protein-10 (BMP-10) encoded by an isolated DNA molecule comprising a sequence selected from the group consisting of:**
> **(a) nucleotides 167 through 1102 of SEQ ID NO: 1;**
> **(b) nucleotides 160 through 1431 of SEQ ID NO: 10;**
> **(c) naturally occurring allelic sequences and degenerative codon sequences of (a) or (b); and,**
> **(d) DNA sequences which hybridize to the complement of (a) or (b) under hybridization conditions of 0.1 times SSC 0.1% SDS at 65° C.**

Example 19 (claim 1 from US Patent No: 5,932,780*) reads:

> **- A transgenic mouse whose genome comprises a transgene comprising:**
> **an AChE promoter operatively linked to a DNA sequence selected from the group consisting of SEQ ID NO:1; SEQ ID NO:3 and SEQ ID NO:5, wherein said sequence is expressed in cells of said transgenic mouse and further wherein said mouse exhibits changes in its neuromuscular junction structure relative to control mice lacking said transgene, wherein said changes comprise longer and more curled postsynaptic folds and a high density of membrane vesicles in the nerve terminals.**

Examples 15 and 18-19 recite claims for which the sequences defined by their "SEQ ID NO:" are not the claimed subject matter. These sequences only help to define the inventions (i.e an antisense compound, an antibody and a transgenic mouse) more precisely.

Example 20 (claim 1 from US Patent No: 6,645,509*) reads:

> **- A composition comprising, in a physiologically acceptable medium, one or more purified natural or synthetic polypeptide(s), wherein each of said polypeptide(s) comprises the amino acid sequence of SEQ ID NO: 1.**

Example 21 (claim 1 from US Patent No: 6,183,968*) reads:

> **- A composition comprising a plurality of polynucleotide probes comprising a nucleotide sequence of *SEQ* ID NOs:1-134.**

Example 21 is a claim for a composition for use in with DNA arrays[74].

A composition of, for example 1489 of these probes would not be protected by Patent US6500938[75]. A DNA microarray consists of a series of DNA elements arranged as spots in a grid pattern on a miniature solid support. These arrayed targets are hybridized with a complex probe comprising many different sequences which is prepared from an RNA population from a particular cell type or tissue. The composition of the probe reflects the abundances of individual transcripts in the source RNA population.

II. Method claims

Typical method claims:

-use of a gene to diagnose disease or disorders associated with the gene,
-us of a gene or the protein it codes for, as a therapeutic to treat a disease or disorders associated with the gene,
-method of identifying molecules that modulate or interact with the gene or the protein it codes for,
-gene therapy.

---

[74] A DNA microarray (also commonly known as gene chip, DNA chip, or biochip) is a collection of microscopic DNA spots attached to a solid surface. Scientists use DNA microarrays to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. (http://en.wikipedia.org/wiki/DNA_microarray).

[75] Supra note 40

Example 22 (claim 1 from US Patent No: 6,737,232*) reads:

- A method of screening for a bioactive agent capable of modulating apoptosis in a mammalian cell, comprising: a) contacting a candidate bioactive agent with a mammalian cell expressing a recombinant nucleic acid encoding an ING2 protein; and b) measuring apoptosis in said mammalian cell; wherein said recombinant nucleic acid comprises a nucleic acid sequence selected from the group consisting of the nucleic acid sequences set forth in SEQ ID Nos: 1, 3, 5, 7, and 9, wherein said ING2 protein will bind to an inhibitor of apoptosis protein (IAP) in the absence of said candidate bioactive agent, and wherein a change in the inhibition of apoptosis indicates that said candidate bioactive agent is capable of modulating apoptosis in a mammalian cell comprising an ING2 protein.

Example 23 (claim 1 from US Patent No: 6,432,644*) reads:

- A method for diagnosing the presence of a polymorphism in human KCNE1 (the coding region of which is bases 193-579 of SEQ ID NO:3) which causes long QT syndrome wherein said method is performed by means which identify the presence of said polymorphism, wherein said polymorphism is one which results in the presence of a KCNE1 polypeptide of SEQ ID NO:4 with an altered amino acid.

Example 24 (claim 1 from US Patent No: 6,440,676*) reads:

- A method for determining the effectiveness of a treatment regimen for brain cancer, the method comprising:

(a) removing a control biological sample from a patient prior to treatment;

(b) determining the amount of transcription in the control biological sample of a nucleic acid transcript selected from the group consisting of SEQ ID NO. 1, SEQ ID NO. 2, SEQ ID NO. 3, SEQ ID NO. 4, SEQ ID NO. 5, SEQ ID NO. 6, SEQ ID NO. 7, SEQ ID NO. 8, and SEQ ID NO.9;

(c) removing a test biological sample from a patient following treatment; and

(d) determining the amount of transcription in the test biological sample of a nucleic acid transcript selected from the group consisting of SEQ ID NO. 1, SEQ ID NO. 2, SEQ ID NO. 3, SEQ ID NO. 4, SEQ ID NO. 5, SEQ ID NO. 6, SEQ ID NO. 7, SEQ ID NO. 8 and SEQ ID NO. 9; wherein a decreased amount of transcription in the test sample relative to the control sample indicates that the treatment regimen for brain cancer is effective.

Example 25 (claim 1 from US Patent No: 6,723,705*) reads:

- A method for treating a mammalian subject having a solid tumor, comprising direct injection of a nucleic acid molecule encoding B7-2 molecule in a form suitable for expression of the B7-2 molecule, into cells of the tumor, wherein the 87-2 molecule has the ability to costimulate a T cell and the ability to bind a CD28 or CTLA4 ligands, such that the growth of the tumor is inhibited.

Example 26 (claim 1 from US Patent No: 6,127,145) reads:

- A method for producing a stable, glycosylated .alpha.1 -antitrypsin (AAT) comprising the steps of:

(a) obtaining rice or barley cells transformed with a chimeric gene having (i) an inducible transcriptional regulatory region, (ii) a first DNA sequence encoding AAT, and (iii) a second DNA sequence encoding a signal polypeptide, where said second DNA sequence is operably linked to said transcriptional regulatory region and said first DNA sequence, and where said signal polypeptide is in translation-frame with said AAT and is effective to facilitate secretion of expressed AAT from the transformed cells, ..