INVENTORSHIP ATTRIBUTION IN ACADEMIC PATENTS: QUANTITATIVE ANALYSIS OF PATENT-PUBLICATION PAIRS[⊗]

Francesco Lissoni[◊✱⌘], Fabio Montobbio [▫✱]

[◊]Università degli studi di Brescia (Italy)
[▫]Università dell'Insubria, Varese (Italy)
[✱]CESPRI – Università "L. Bocconi", Milan (Italy)

[⌘] *corresponding author: francesco.lissoni@unibocconi.it*

2nd draft, June 2006

1. Introduction

Recent research in economics and business has highlighted the increasing importance of patenting activity by universities and individual scientists, both in the US (Henderson et al., 1998; Mowery et al., 2004), and in a few European countries (Meyer et al., 2003; Balconi et al., 2004; Gering and Schmoch, 2003).

Issues of related to the ownership of academic inventions have attracted a good deal of attention, whether they have to do with the so-called "anti-common" problem (for a survey: Stephan, 1996) or the social efficiency of patent regimes which contemplates the so-called "professor's priviledge", which assigns IPRs over academic research results to the scientists who got those results, and not to the universities wherein the scientists are employed (PVA-MV, 2003; OECD, 2003, ch.1).

Compared to the ownership issues, however, much less attention has been paid to problems in the attribution of inventorship. In particular, it has not been explored what criteria teams of scientists follow when it comes to turn the results of collective research efforts into one or more patents, and to decide who will be credited as inventor. A cursory look to both patent and publication data, in fact, reveals immediately that the average number of authors of scientific papers is higher than the average number of inventors listed on patents, even when the latter come from a science-based field, possibly one that is dominated by academic research (Ducor, 2000). Some scientist, credited as author of a research-related paper, is left out from the patent. Who? And why?

To the extent that the largest part of the academic research leading to patents is the outcome of collaborative research efforts, the problem of inventorship attribution is a delicate one. All national IPR legislation contemplates a number of rights and benefits for inventors, which the patent owner is bound

---

[⊗] We thank Maurizio Tosetti for his invaluable research assistance. A preliminary draft of the paper was presented "Con/texts of Invention" conference, Case Western Reserve University. Cleveland, Ohio April 20-23, 2006

to respect. The US legislation, in addition, states that the exclusion of one or more legitimate inventors from the patent document, although amendable, may ultimately lead to the invalidation of any granted patent (Fasse, 1992; pp.73-74).

Problems in the attribution of inventorship over academic research may result into social inefficiencies for at least two reasons.

1. Even when they do not lead to the invalidation of the patent, conflicts on inventorship attribution may lead to court litigation, which is a key and growing component of the social costs of the patent system
2. Patents are increasingly counted alongside with publications as indicators of academic scientists' productivity, especially by governmental assessment exercises. As a consequence, misplaced inventorship claims could detract from, or boost undeservedly, a scientist's reputation, and multiply existing distortions in the economics of academic careers, such as the Matthew effect (Merton, 1988).

In this paper, we try to measure the extent and nature of the above-mentioned problems, by comparing patterns of co-inventorship and co-authorship. First, for a large number of Italian "academic inventors" (academic scientists designated as inventors on patents owned either by their universities or research sponsors) we identify so-called patent-publication pairs, i.e. patents whose contents have also been divulged through scientific publications (or, to put it in another way, publications whose contents have also been patented). We build different samples according to different methodologies for the identification of the patent-publication pairs, which cover four distinct disciplinary fields. Second, we give a quantitative account of the difference between the number of authors and inventors per patent-publication pair. Last, we investigate the academic status of the scientists listed as authors of the publications, but not designated as inventors of the patents. To the extent that the academic status, and not the authors' effective contribution to the patented inventions, is the cause for exclusion, we envisage a potential for conflict. Our exercise, however, sheds light not only on inventorship attribution problem, but also on the long-standing problem of authorship attribution in science. By comparing criteria of attribution of authorship and inventorship one can get insights on issues such as gift authorship, and the requirements imposed on scientists in order to qualify as authors of the papers they contribute to.

The paper is structured as follows. In section 2 we introduce the concepts of co-inventorship and co-authorship, as they are presented, respectively, in the legal and sociological literature, as well as in "grey" literature of journals' publication guidelines and technology transfer offices' recommendations to potential academic inventors. In section 3 we describe our data, and delve into some methodological details regarding our publication-patent matching exercise. In section 4, we report our evidence on the difference between the number of authors and inventors in selected patent-publication pairs in four different disciplinary fields and our results on authors' exclusion from patents and on the identity of the excluded authors. In section 5 we draw some conclusions and set the stage for further research.

2. Co-authorship and co-inventorship

After growing incessantly for over half a century, multiple-authored publications now dominate many scientific and technical fields, while the average number of authors per publication keeps increasing (Weeks et al., 2004). Historians and sociologists of science have explained this trend with the changing nature of the scientific work, which has been placing increasing emphasis on collaboration, inter-disciplinarity, and the sharing of data and facilities (Katz and Martin, 1997).

Many scholars have argued that, with the loss of correspondence between individual papers and authors, the concept of authorship itself has become an increasingly fuzzy one, which is better explained by social and legal conventions, than by a "natural" identification of the scientist with her own work and output (McSherry, 2001; Galison and Biagioli, 2003). In the medical field, for example, journal editors trust the definition of authorship to a complex set of rules, which allow for a wide heterogeneity in the relative size and nature of the individual scientist's contribution to a co-authored paper. According to guidelines of the *International Committee of Medical Journal Editors*, a scientist who has just "conceived and designed" a paper, limited himself to a "critical" revision of the manuscript, and obtained the necessary research funds, can qualify as an author, alongside with a colleague who has acquired, analysed and interpreted the data, drafted the manuscript, and provided statistical expertise (ICMJE, 2006). Similar rules, albeit less detailed, can be found in the authors' guidelines of the International Electrical and Electronic Engineering association (IEEE, 2006).

Compared to co-authorship, co-inventorship has been much less investigated. A cursory look at the time series of European or US Patent Office data, however, would reveal that co-inventorship has been on the rise for quite a few years now [1].

At the same time, comparisons with publication data reveal that the average number of inventors per patent is well below the average number of authors per patent, even for comparable technological and scientific fields (Meyer and Bhattacharya, 2004). One possible explanation for this difference is that while publications are the realm of academics, patents derive mostly by industrial research funded by business companies and performed by those companies' employees: the proprietary nature of this research forces caution in looking for cross-firm collaboration, and in granting to industrial researchers the same freedom of choosing research team partners enjoyed by academic scientists.

---

[1] The average number of inventor per patent applications at EPO (the European Patent Office) has increased constantly, from 1,95 in 1980 to 2,46 in 1999. When considering only patents in science-based field such as organic chemistry, the figures are respectively 2,76 and 3,88. (Source: authors' elaboration on EP-CESPRI database).

However, differences in the number of co-authors and co-inventors are found also for so-called "patent-publication pairs", that is science-based inventions which come from the same research team and are both patented and described in publications (Ducor, 2000; Murray, 2004; Murray and Stern, 2004). In this case, the smaller number of inventors must be explained by the exclusion of some authors of the publication from the set of designated inventors of the patent. In turn, the most immediate explanation for such exclusion is that the criteria according to which inventorship is granted are more restrictive than those defining authorship.

Indeed this is the explanation most frequently put forward by the many consultants and university technology liaison officers who dispense advice to academic scientists looking for IPR protection (Bennett and Biswas, 1997; Vinarov, 2003; Hutchins, 2003). These experts argue that the criteria for qualifying as an inventor are more restrictive because the inventorship concept is defined by the law, as opposed to the (lax) social conventions to which the definition of authorship is trusted. In the words of an employee of the US Patent Office, the typical argument is that:

> … inventorship is very unlikely authorship in the nonpatent literature. Inventorship […] is a very precise legal concept which identifies the developer of the patented technology. Only a substantial contributor to the technology can be an inventor. Authorship, on the other hand, often is a matter of convention or the preference of the principal author. A common practice today is to include as coauthors of technical papers those who contributed in only a minor way to its development (Terapane, 1977; p. 130)

Digging into the law literature, however, reveals that the concept of inventorship is much more controversial than it may appear from such sweeping statements. In particular, the US literature offers a few enlightening discussions of what has been termed "the muddy metaphysics of joint inventorship", and how this may be used to explain the exclusion of authors from inventorship (Fasse, 1992)[2]. These discussions suggest that no clear criteria are ready at hand to decide which authors do not qualify as inventors, at least on the sheer basis of the authors' research contribution.

2.1 Mud-racking: reasons for authors' exclusion from inventorship

According to section 35 of the US constitution (as amended in 1984), two individuals can be named as inventors on the same patents only if they have worked "jointly" and provided some kind of "inventive" contribution.

Of these two requirements, "joint manner" and inventiveness, only the latter can be responsible for the eventual exclusion of an author of a scientific paper from the related patent. In fact, very much like co-authors, co-inventors may qualify as such even if they do not work together (at the same time or in

---

[2] Co-inventorship issues are much more relevant for the US patent system, which grants patents on a first-to-invent basis, than for the patent systems of other nations, which are usually based upon the first-to-file principle. According to the US law, mistakes or omissions in the designation of inventors can be corrected by the inventors themselves or by the court, but can also cause the invalidation of the patent (Fasse, 1992; pp. 172-173)

the same place) and do not contribute to the invention in the same way. For example, in a multi-claim patent they may be responsible both for different claims and for a disproportionate number of claims; or they can have undertaken different, non-overlapping research tasks.

On the contrary, contributing with some inventive conceptual contribution to the invention is a requirement for inventorship that some authors of scientific publications may fail [3]. However, it is hard to tell when a contribution is not inventive, nor "conceptual" enough, and whether this is the real cause of exclusion in common practice.

If we consider a typical scientific paper, the authors listed in the by-line will include both graduate students and junior scientists, and possibly laboratory technicians, alongside with senior figures such as the students' supervisors or the laboratory heads.

To the extent that current interpretations of the law suggest that "merely suggesting a desired result" or "having entrepreneurial involvement" do not qualify as inventorship[4], we may expect the most senior authors of a paper to be at risk of exclusion from the patent, as long as their contribution is limited to raising funds, conceive the initial experiment, and revising the draft paper (which would qualify them as authors, at least according to the ICMJE guidelines). In fact, the position of some laboratory heads or supervisors may be very close to that of an entrepreneur to the extent that, like him, they pose a problem to be solved, hire and manage human resources, suggest a desired result, refuse a few initial solutions provided by the employees, and finally approve the most satisfactory result. All these acts together do not qualify the entrepreneur for inventorship, but they give him substantial claims over ownership. However senior "entrepreneurial" scientists are very likely to be deprived of the latter by their university or research sponsors. Accordingly, they may feel that being credited with inventorship is a form of compensation, which can be formally justified just by stretching to some extent the economic logic according to which they already reclaimed authorship [5].

---

[3] Indeed, judicial cases reported by Fasse (1992, p.199) mention explicitly co-authorship as an insufficient test for inventorship

[4] Fasse, 1992; pp. 192ff.

[5] The economic logic behind rewarding through authorship a scientific entrepreneur, such as a laboratory head, is well stated in Carl Djerassi's fictionalized portrait of a senior scientist explaining why she deserves co-authorship along with her junior partner:

> "I am the one who suggested the problem […]. I provided the facilities and [the junior scientist's] fellowship through my research grant. I prepared the grant application to the NIH. In it, I outlined in great detail what my research group was going to do, why it is important, what the earlier contributions were, and many other things. Without such support [the junior scientist] could do nothing. […] I'm talking about all the instruments in my lab, the chemicals, the glassware." (Djerassi C., 1989; pp. 50-51).

To be fair, Djerassi's senior scientist adds that she contributes also intellectually, but we may question whether it is enough for being considered either an inventor or an author:

> "… I see [the junior partner] almost every day; we discuss the progress of the work; I suggest certain techniques; I call important references to her attention […] there's both a teacher-apprentice relationship and collegiality" (Djerassi C., 1989; p.51).

At the opposite end, “following the complete instructions” of a colleague or superior does not qualify a researcher as an inventor; and even joining too late a research team, whose members have already conceived the key characteristics of the desired invention, may be a cause for exclusion from inventorship. These cases remind naturally to borderline situations in which a junior scientist or a graduate student whose brilliant assistantship (which often requires creativity) has been rewarded with authorship, but not with inventorship [6].

It is possible that credit for authorship and inventorship depends upon the discretionary judgement of the most senior scientists, who most often manage the economic details of the research enterprise and exercise authority, and whose opinion may carry an heavy weight even in group’s decisions. When faced with the difficult task of evaluating their junior colleagues’ contribution towards obtaining some research results, these authors may be tempted to stretch their judgement in a favourable direction when confronted with the problem of the attribution of authorship (which entails only a reputational reward), and in the opposite direction when deciding upon inventorship (which may also lead to more tangible economic benefits). In doing so, senior authors may be affected by a tendency to overvalue their own contribution, both to patents and publications, which questionnaire surveys by Hoen et al. (1998) and Jaffe et al. (2000) have shown to be a frequent occurrence.

The practicalities of both inventorship and authorship attribution may leave room both for this kind of “stretching” and possibly for more arbitrary acts. Both journal editors and patent office examiners trust entirely the identification of legitimate authors and inventors to the individuals who submit their manuscripts or applications. At most, signed declarations are required. If not challenged in court (either by some excluded individuals, or by one ore more included ones, who contest the unfair inclusion of others) this initial attribution remains un-scrutinized: both the journal editors and the patent examiners, in fact, pay attention only to the technical contents of the papers and patents they are called to judge[7]. It is doubtful that a junior author, excluded from a patent, will find it convenient to suit a senior co-author, upon whom her career prospects may depend heavily. Similarly, we cannot expect the same junior scientist, who has signed both a paper and a patent, to oppose a legal action against the inclusion of an illegitimate author in the publication.

2.2 Proposed analysis

Summing up the previous discussion, we may expect different categories of authors to be excluded from inventorship.

On Djerassi’s authority as a source of information of authorship practices in science, see Garfield E. (1983). For non-fictional, quantitative accounts of “honorary” authorship see Hoen et al. (1998) and Mowatt et al. (2002), as well as references therein.

[6] For a case of a student’s exclusion from a patent, see Fasse, 1992; p. 282

[7] Editors will check, through the referees, the relevance, originality and rigour of the paper. Examiners will check for novelty, non obviousness, and industrial applicability

I. Senior scientists whose entrepreneurial contribution to the research enterprise qualify them for authorship, accordingly to guidelines such as the ICMJE's, but not for inventorship, according to the rule of law

II. Laboratory technicians and other assistant figures (including graduate students and junior scientists in charge of minor tasks), who have been rewarded with "honorary" authorship, but do not qualify for inventorship

III. Junior scientists, whose practical task has been creative enough to qualify them for authorship, but not for inventorship (this includes the case of junior scientists being involved in the team at a later research stage)

IV. Junior scientists who qualify both for authorship and inventorship, but are arbitrarily excluded from the latter for economic reasons.

Of course these four categories are too sharp to describe all possible situations. In between the discrete categories of junior and senior scientists there is a continuum of scientists with different degrees of seniority. Similarly, what is "arbitrary" and what is "honorary" is a matter of viewpoints, and hardly measurable.

However, the four categories lend themselves to some form of identification through bibliometric data.

In particular, we may think of exploiting the information provided by the order according to which authors are listed in the by-line of scientific publications. Although authorship guidelines, such as the ICMEJ's or the IEEE's do not provide explicit recommendations, two major traditions exist in this respect: alphabetical ordering (which is typical, for example, of social sciences) and contribution-related ordering, which is most common in the hard science. Seniority-based criteria are also common, according to which senior authors are listed last, which in fact is very near to admitting that he or she contributed the least[8].

Therefore, in the case of non-alphabetical by-lines of scientific papers (whose contents have also been patented) it is reasonable to presume that first authors are both those who contributed most to the paper and are most likely to qualify for inventorship. At the opposite end, last authors are very likely to be senior scientists whose contribution to the publication mostly resembles the kind of "entrepreneurial involvement" that may or may not qualify them for inventorship, depending on the additional conceptual contribution they may have added.

---

[8] In their study on medical publications, Mowatt et al. (2002) calculate that 76% of by-lines list the person contributing primarily to the study first, while only 2% list authors alphabetically. Of the remaining 22%, seniority criteria were involved, such as listing the senior author last. Some professional societies explicitly recommend to list authors according to their contribution.

Accordingly, whenever we find that the first author of a non-alphabetical paper has been excluded, we may suspect we are facing category *IV.* of those listed above.

Exclusion of last authors, on the contrary, may be indicative of the relevance of category *I.*

As for categories *II.* and *III.*, they may lead to the exclusion of authors listed somewhere between the first and last position of the by-line.

The first results of this kind of analysis, mainly descriptive statistics, are presented below, in sections 3 and 4.

3. Data and Methodology.

This section is devoted to the description of the data and methodology. Our data result from the integration of three different sources: the EP-INV-DOC database on Italian academic inventors, their publication record according to the ISI Science Citation Index, and a few extra information obtained by a sample of academic inventors through a questionnaire. In section 3.1 we say more on these sources, while in section 3.2 we explain the text mining methods we used to select the patent-publication pairs in four different disciplinary fields.

3.1 Description of the Sample

The EP-INV-DOC database contains all the 919 active Italian academic scientists designated as inventors on patents owned either by universities, public research organizations or business companies, both in Italy and abroad. It contains also information on individual characteristics of the scientists (such as age, affiliation, academic rank, discipline) and the information that can be found in the front page of their patents (including title and abstract). It originates from the complete list (provided by the Italian Ministry of Education) of professors and researchers who, in 2000, held a position in a scientific or technical discipline in an Italian university (including medical and engineering schools). Names and surnames from that list were matched to names and surnames in the EP-INV database, a broader dataset which contains all patent applications to the European Patent Office which designate at least one inventor with an Italian address, from 1978 to early 2000[9].

In the EP-INV-DOC database, disciplines are defined according to a classification created by the Ministry for administrative purposes, which is very detailed and allows some compression into broader

---

[9] Overall, the EP-INV database contains information on 30,243 inventors and 38,868 patent applications (for a more comprehensive description, see Balconi et al. 2004). The original list of professors contains little more than 30,000 names. For sake of simplicity, we will refer to patent applications simply as ‘patents’.

categories, which we will refer to as 'fields'[10]. We focus on the four fields with the highest share of academic inventors over the total number of professors in the field, namely Chemical Engineering (it includes technology of materials, such as macromolecular compounds), Biology, Pharmacology, and Electronics and Telecommunications, for a total of 308 academic inventors and 552 patents.

The major limitation of the EP-INV-DOC database, is that it includes only the scientists who have passed a competitive examination for a tenured position (from now on, we will refer to them simply as 'professors')[11]. Thus our data miss the large number of researcher, at the time of the patent or publication, were working as fixed-term appointees or PhD students, and did not manage to get a tenure before 2000. We also miss university technicians.

Publication data were collected from the 1975-2003 on-line version of ISI *Science Citation Index* for all the 308 of the academic inventors in the selected fields. A more detailed description of this initial sample can be found in Breschi et al. (2005 and 2007). Finally we sent a questionnaire to a sub-sample of 110 scientists, asking questions related to the history of each patent, including whether the scientific research behind the patented invention had also produced one or more academic publications.

3.2 Methodology

Our methodology is based on the identification of *patent-publication pairs*. A patent and a paper form a pair when the same idea is described in both documents. This happens when a new scientific idea coincide with a solution to a technical problem and has some degree of industrial applicability.

Scientific papers and patents differ widely in contents, since scientific publications describe a set of theories and/or experimental results, of which they emphasize the originality and neatness according to some rhetorical rules, while patents describe the features of a new product or process, of which they emphasize the novelty and utility, by laying out a list of claims. However, in so-called "science-based" technologies and in engineering, it is often the case that a patentable advancement is also worth of publication in refereed journals. In this case, we may expect highly specific words to be present both in the patent and the publication that report on the advancement; inventors and their lawyer may also cut-and-paste a few sentences from one document into another.

---

[10] The major limitation of the MIUR list and, as a consequence, of the EP-INV-DOC database, is that it includes only those professors and researchers who had passed a competitive examination for a tenured position (from now on, we will refer to them simply as 'professors'). Thus our data miss the large number of fixed-term appointees who, at the time, had been working in one or more universities for one or more years, as well as all the PhD students, post-doc fellows, and technicians. In the current Italian system , assistant professor (called 'researcher') and associate professor positions, despite being only the first two steps of the academic career, are not offered as fixed-term appointments, but as tenured ones. The main differences with the position of full professor lie in wage and administrative power.

[11] In the current Italian system , assistant professor (called 'researcher') and associate professor positions, despite being only the first two steps of the academic career, are not offered as fixed-term appointments, but as tenured ones. The main differences with the position of full professor lie in wage and administrative power.

To our knowledge only Ducor (2000), Murray (2002) and Murray and Stern (2005) have tried to build patent-publication pairs. In order to do so, they have used different methodologies. Ducor (2000) performed a manual search of various databases for proteins with specific genetic or aminoacid sequences and found 40 article-patent pairs. Ducor himself recognises that his data are too few and biased to be statistically representative. Murray's (2002) study concerns only a single patent-paper pair on tissue engineering in cartilage. Murray and Stern (2005) have examined 340 articles published on Nature Biotechnology between 1997 and 1999, and the US patents invented by the authors of the articles, or assigned to their institutions; finally, through careful reading, they selected 169 patent-publication pairs.

In this paper we use a different methodology, as we try to lay down general methodology to build larger sets of patent-paper pairs in different technological fields. In particular, we rely upon established quantitative methods of data mining and information retrieval, together with an analysis based on our questionnaire.

We start from all Italian academic inventors in the four disciplines we mentioned above. Then, we couple their patents to all their publications spanning two years before and after the priority date of the patents and create a set of *potential* patent-publication pairs. From this initial set, we select the *actual* patent-publication pairs according to five different methods, three of which are based on the extraction of information from textual data, in particular from titles and abstracts of patents and publications (Leopold et al. 2004; Bassecoulard and Zitt, 2004), one is based on our questionnaire, and one on the combination of the previous four.

3.3 The selection of the patent-publication pairs

In order to find out the relevant patent-publication pairs we start from all the patents signed by the individual scientists in the sample. Given t the priority year of a patent and i the individual listed among the inventors, a *potential* patent-publication pair is defined as the association between the patent and a publication that has individual i among the authors and has been published in the period $[t-2, t+2]$.

Excluding duplications (which may occur when two or more patents or two or more publications have the same co-inventors or co-authors), all publications with no abstracts, and all patents for which the inventors declared in a questionnaire that there are no related publications, the final sample of *potential* patent-publication pairs is composed of 6810 pairs, 218 individuals, 389 patents and 2838 publications.

A description of this sample is provided in Tables 1, 2, 3 and in Figure 1. Table 1 shows the number of academic inventors by field and years of birth. Table 2 shows the number of patents by priority date and type of applicant. It's worthwhile noting that 63% of patents are owned by a business company[12]. Table 3 shows the number of publications by year of publication and fields of activity of the authors. 60% of the publications in the sample are from scientists in the fields of Biology and Pharmacology.

Figure 1 shows for the 2838 publications and 389 patents the observed frequencies of the number of authors in each publication and of the number of inventors in each patent. The distribution of the number of authors has a fatter tail to the right. The median number of authors per publication is between 4 and 5. The median number of inventors per patent is between 2 and 3. Moreover the maximum number of inventors per patent is 21 and there are 23 publications with a number of authors greater than 21 (the top two publications have 337 and 517 authors respectively).

From this initial set of *potential* patent-publication pairs, we select the *actual* patent-publication pairs using five different methods. Three of these methods are based on the extraction of information from textual data, in particular from titles and abstracts of patents and publications (Leopold et al. 2004; Bassecoulard and Zitt, 2004). We want to exploit in a systematic way the lexical linkages between a patent and a publication and to give a quantitative appraisal of their similarity using the words contained in the documents. Therefore the first step is to remove the uninformative words (so called *stop words*) from titles and abstracts. These very frequent terms are equally distributed across titles and abstracts, and are not relevant for the identification of the patent-publication pair. Stop words include pronouns, conjunctions, and the most frequent nouns and verbs. After the exclusion of stop words we proceed as follows:

METHOD 1 (M1): For each potential patent-publication pair we compare the patent and publication abstracts, and calculate the number of words that are the same in the two documents. Then we calculate the share of words that are the same on the total number of words in the patent abstract. We then select the patent-publication pairs that are on the top 10 percentile with a floor value of the share equal to 11,86.

METHOD 2 (M2): This method selects the patent-publication pairs simply on the basis of the answers to the questionnaire, filed to a subset of 110 academic inventors, which we have described above. In particular, we retain as actual patent-publication pairs only those by academic inventors who

---

[12] Most patents belong to business companies, as a result of contractual funding, with little meaningful differences across fields. This percentage is slightly lower than in Balconi et al. (2004) and in Breschi et al. (2007).This is because industry owned patents have an higher probability to be excluded in the selection of the potential patent-publication pairs. IPRs over public-funded research, in principle these belong to the MIUR ministry, the National Research Council, and, in the past, ENEA, the National Agency for Alternative Energy. However, until recently, the decision to take the first step towards patenting was usually left to grant recipients. Moreover until recently, universities decided to take charge of the application procedure and expenses more to reward, often symbolically, some brilliant researcher, rather than as the outcome of a consistent exploitation strategy. As a result, few patent applications from public-funded research are completed. It also happens that many professors take the shortcut of patenting in their own names: this explain the presence of a few inventors' own patents.

confirmed that the scientific research leading to the patented invention also produced technical or scientific publications. The total number of patent-publication pairs in this case is 3380.

METHOD 3 (M3): M3 is based on a traditional data mining method called *bag of words* which the literature considers simple and effective (Leopold et al. 2004, Salton and McGill, 1983). First we divide the potential patent-publication pairs in four distinct groups according to the four disciplinary fields to which individual scientists belong. For each disciplinary field we build a complete set of words from all titles and abstracts of the patents and publications (with he exclusions of the stop words). Each document (patent or publication) j can be represented by a vector. Each cell (i,j) in the vector has a value equal to 1 if the word i appears in document j and 0 otherwise (Bassecoulart and Zitt, 2004). This vector representation may be used to produce a large number of similarity measures between patents and publications. The most common one, which we use here, is the cosine similarity measure (S). If x_{ij} is the value of the binary variable for document j and word i, S measures the similarity between a document k and s as follows:

$$S(k,s)=\sum_i x_{ki}x_{si}\Big/\sqrt{\sum_i x^2{}_{ki}}\sqrt{\sum_i x^2{}_{si}}$$

Finally, as for M1, we then select the patent-publication pairs on the top 10 percentile, whose S value range, in our case, from 0.145 to 0.75 (S theoretical values range from 0 to 1).

METHOD 4 (M4): It is equivalent to M3, with the only difference being that cells in the vectors do not contain dummies but frequencies, that is the number of occurrences for each word in the documents. After calculating S with this type of vectors, we select once again the patent-publication pairs on the top 10 percentile (with S ranging from 0.206 to 0.81).

METHOD 5 (M5): It is the intersection of the 4 methods listed so far: only pairs that meet criteria set by all methods are retained. It results in a sample of 104 patent-publication pairs.

The analysis of the selected samples shows that the concept of patent-publication pair can be difficult to turn into an empirically viable measure. Even when focusing only on the 104 pairs from M5, multiple relationships patents and publications can be observed. We have neat "one-to-one" relationships (one patent corresponds to just one publication, and *vice versa*) only in 23 pairs out of 104. No less than 44 pairs are the result of "one-to-many" matches between one patent and many (n) publications (more precisely, we have 13 patents that originate 44 pairs by combining with as many publications). We also have 9 patents and 9 publications that are all interconnected in different ways ("many-to-many" matches) and originate 36 patent-publication pairs.

Compared to the number of "one-to-many" and "many-to-many" occurrences, the "many(patents)-to-one(publication)" case is much rarer (we have only two publications that are associated with two patents each, for a total of four pairs). This can be explained with the common-sense intuition that a good research project will produce more than result worth of publication. Although published separately (in order to keep the length of articles under control, or simply to follow a "salami slicing" strategy) these results are likely to be patented jointly, since patent system provides many incentives to pool many claims in a single application. First, there are cost incentives: application and renewal fees are set per patent, not per claim. Second, broader patents are more effective in protecting the economic exploitation the invention.

The quantitative relevance of the patent-publication pairs originated by the one-to-many and many-to-many matches bears two important methodological consequences.

First, it suggests that, when it comes to checking which authors of a paper are excluded from inventorship, the most appropriate unit of analysis may not be the list of authors(inventors) provided separately by each publication(patent), but the overall team of authors(inventors) listed in the set of related publications(patents).

Second, we observe that within these teams a specific division of labour may be implemented, which may be responsible for the specific patterns of co-authorship and co-inventorship we observe.

Accordingly, focusing on the single patent-publication pairs may be misleading, since our data mining techniques are not able to select pairs that are absolutely identical in terms of the ideas they contain. Therefore we present our results after having aggregated authors and inventors in specific teams built on the basis of the links between patents and publications in related pairs.

4. Results

We present two types of results. In section 4.1 we document the extent of the author-inventor bias for the five different methods of pair selection and for the four different disciplinary fields; that is, we give a quantitative account of the difference between the number of authors and inventors in comparison with the total sample of potential patent-publication pairs. In section 4.2, we analyse the patterns of authors' exclusion. In order to do so, we take into account only the publications whose authors are not listed in alphabetical order, and look at the position in the by-line of the excluded authors.

4.1 Number of authors and number of inventors

4.1.1 Results for the five different methods of pair selection

In Panel 1 we show the distribution of the difference between the number of authors and the number of inventors across five different sets of patent-publication pairs, one for each of the five selection methods. The number of patent-publication pairs are 681 for M1, M3 and M4, 3380 for M2, and 104 for M5. In Table 4 we present the average number of authors and inventors and the average and median difference between the number of authors and inventors. Panel 2 and Table 4 show also the results for the four disciplinary fields.

The number of inventors of a patent is lower than the number of authors of the publication that contains the same idea. The average number of inventors across the patent-publication pairs is on average between three and four, while the number of authors is around five. It is interesting to note that, according to M3 and M4 (i.e. the quantitative data mining techniques) and to the further selection of M5, the average and median differences between authors and inventors are reduced and the distributions displayed in Panel 1 are more concentred around the mean, which in turn is closer to one (for M3, M4 and M5 the median and average difference are closer).

Looking at five selected samples of patent-publication pairs we find evidence of a process of exclusion. At the same time the size of this effect is inferior to the one in the general sample when patents and publications are compared without a textual analysis of their titles and abstracts (e.g. from Figure 1 or from line 1 in Table 4).

Finally, focusing only on the 104 pairs on the intersection of the four selected samples (M5), multiple observations for patents and publications can be detected as explained in the previous section. In particular excluding the 23 pairs with a one to one relation between patents and publications, the remaining pairs can be collapsed in 19 groups of patent-publication pairs. Each group contains a set of pairs that are linked among each other because one patent can be associated to more than one publication and *vice versa*. So we decided to count in each group the total number of authors and inventors to see whether an exclusion pattern at the group level can be detected. The final graph (*collapsed sample*) in Panel 1 shows that on average the number of authors in each team remains higher than the number of inventors. So even if the sets of patents and publications are the result of a team, somebody in the team is likely to be excluded from the benefits of inventorship.

4.1.2 Differences across disciplines

Significant differences across disciplines emerge. We show the result just for M3 and we have 178 in Biology, 201 in Pharmacology, 62 in Chemicals and Material Technology and, finally, 240 in Electronics and Telecom. Panel 2 and Table 4 show that there is a significant difference between the number of authors and the number of inventors only in Biology and Pharmacology where there is an higher average number of authors. At the same time the difference is very small in Chemicals and Material Technology and Electronics and Telecommunications. In this two disciplines we find approximately the same average number of authors and inventors and the median value of the difference across pairs is equal to 0. It is also interesting to note that in our sample Chemical and Electronics are the two sectors in which there is a larger share university patents owned by private companies and Biology and Pharmacology have conversely a relatively higher share of patent owned by universities and individual scientists. This may suggest that the decisions concerning inventorship may be associated with the contractual norms that determine the ownership of the patent.

4.2 Patterns of Exclusion

For our five samples of patent-publication pairs we investigate whether a specific pattern of exclusion emerges. We select those pairs in which the authors of the publications are *not* listed in alphabetically order and with the number of authors greater or equal to the number of inventors. We order the publications by number of authors and then single out the excluded author's position in the by-line.

Our assumptions are that first authors are both those who contributed most to the paper and, at the opposite end, last authors are very likely to be senior scientists, whose role has been mainly of an "entrepreneurial" kind. Results for the five methods - displayed from Table 5 to 9 - show that count of exclusions by position of the excluded author in the by-line and number of authors per publication. For sake of simplicity, but without loss of relevant information, we include the publications only up to 14 authors.

All the tables exhibit the same pattern of exclusion. They indicate that the last authors listed in the by-line of the paper title, who we may presume to be the senior scientist heading the research team, have the lowest probability to be excluded. Therefore, contrarily to hypotheses I. outlined in section 2.2, we find that senior scientists (who may have given a limited contribution to the invention) are unlikely to be excluded from inventorship. The second and the third authors in by-line are excluded relatively more often; but also for the first author we find a non negligible number of exclusions.

When we considered the four disciplinary fields separately, we do not detect any significant difference across fields[13].

In conclusion, the descriptive analysis we have produced so far shows that the laboratory technicians and other assistant figures (including graduate students and junior scientists), that we expect to find listed neither in the first nor in the last position, have the highest probability of being left out from inventorship, even if rewarded with authorship (as suggested by hypotheses II and III in section 2.2).

---

[13] Results are not displayed but are available from authors

However our work does not exclude that anyone from the possible categories of authors outlined in section 2,2 may be left out from inventorship. In particular since in many cases also the first author is left out there could be less frequent circumstances in which a junior scientist or the scientist giving the principal contribution to the paper is not listed among the inventors (as suggested by hypothesis IV).

5. Conclusions and further research

In this paper, we have compares patterns of co-inventorship and co-authorship. W have selected five different samples of patent-publication pairs by Italian academic inventors, that is are patents and publications that contain the same new ideas. We have used different methods of pair selection and extended the analysis to four different disciplinary fields. In this way we have given a quantitative account of the differences between the number of authors and inventors in the selected patent-publication pairs. We have also investigated the academic status of the scientists listed as authors of the publications, but not designated as inventors of the patents. To the extent that the academic status, and not the authors' effective contribution to the patented inventions, may be the cause for exclusion, we envisage a potential for conflict. On the other hand, to the extent that inventorship may be a stricter test of substantive research contribution, our results may point to the existence of sizable number of researchers rewarded with "gift" or "honorary" authorship beyond their merits.

Our main result is that the number of inventors of a patent is lower than the number of authors of the publication that contains the same idea. Using quantitative data mining techniques to select the patent-publication pairs we find that the average and median differences between authors and inventors are reduced relatively to the initial sample. Therefore we find evidence of a process of exclusion. At the same time the size of this effect is inferior to the one in the general sample, and presumed by other studies which do not make use of patent-publication pairs.

From the methodological viewpoint, we find that it is far from easy to single out univocally patent-publication pairs, because the patentable and publishable units may be different. We often find a patent related to many publications and many patents related to many publications. This suggests further enquiry on how co-inventorship and co-authorship are negotiated within teams and research projects. And it also suggested caution in accepting results based upon sheer one-to-one patent-publication pairs.

We also find significant differences across disciplinary fields. A significant difference between the number of authors and the number of inventors is detected only in Biology and Pharmacology while in Chemicals and Material Technology and Electronics and Telecommunications is negligible.

Finally a clear pattern of exclusion emerges. The last authors listed in the by-line have the lowest probability to be excluded. This author is typically the head of the department or the leader of the research project and the one that in many cases has given a relatively small contribution to the scientific publication. The second and the third authors in by-line are excluded relatively more often are. However also for the first author we find a non negligible number of exclusions.

In the future drafts of this paper, we will complete our analysis by checking the CVs and publication records of a sample of included and excluded authors. We point at estimating in a robust statistical setting the probability of exclusion of each author and test whether a specific position in the by-line has a statistically positive impact on that probability. In particular we intend to control for a set of individual characteristics that may shed light on the seniority and status (student, technician, postdoc…) of the different authors at the time of the publication and of the patent.

References

Balconi M., Breschi S., Lissoni F. (2004), "Networks of inventors and the role of academia: an exploration of Italian patent data", *Research Policy* 33/1, 2004, pp. 127-145

Bassecoulart E., Zitt, M. (2004); Patents and Publications. The Lexical Connection. In Moed H.F., Glänzel W., Schmoch U., Handbook of Quantitative Science and Technology Research. Kluwer. Dordrecht, Ch. 30.

Bennett V.C., Biswas S.J. (1997), "Protecting the patentability of your collaborative research", *Nature Biotechnology* 15, pp. 472-473

Biagioli M., Galison P. (eds.) (2003) *Scientific Authorship: Credit and Intellectual Property in Science*, Routledge

Breschi S., Lissoni F., Montobbio F. (2005d). "Open Science and University Patenting: A Bibliometric Approach", in Van Pottelsberghe de la Potterie B., De Meyer A. (eds.), Economic and Management Perspectives on Intellectual Property Rights, Palgrave McMillan

Breschi S., Lissoni F., Montobbio F. (2007), The scientific productivity of academic inventors: new evidence from Italian data, *Economics of Innovation and New Technology*, forthcoming.

Djerassi C. (1989), *Cantor's Dilemma*, Penguin Books

Ducor P. (2000), "Coauthorship and Coinventorship", *Science* 289, pp.873-875

Fasse W.F. (1992), "The Muddy Metaphysics of Joint Invetorship: Cleaning Up after the 1984 Amendments to 35 U.S.C. § 116", *Harvard Journal of Law and Technology* 5, pp.73-74

Garfield E. (1983), "Carl Djerassi: Chemist and Entrepreneur", *Chemtech* 13, pp. 534-538

Gering T., Schmoch U. (2003), "Management of Intellectual Assets by German Public Research Organisations", in: OECD, *Turning Science into Business: Patenting and Licensing at Public Research Organisations*, Organization for Economic Co-operation and Development, Paris

HendersonR., Jaffe A.B., Trajtenberg M. (1998) "Universities as a source of commercial technology: A detailed analysis of University Patenting, 1965-1988", *Review of Economics and Statistics* 80, pp.119-127.

Hoen W.P., Henk C.W., Overbeke A.J.P.M. (1998), "What Are the Factors Determining Authorship and the Order ot the Authors' Names?", *Journal of Amercian Medical Association* 280,pp. 217-218

Hutchins M. (2003), "Common mistakes that undermine patent protection and how to avoid them", *International Journal of Medical Marketing* 3, pp. 204-211

ICMJE (2006), *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication*, International Committee of Medical Journal Editors

IEEE (2006) ….

Jaffe A. B., Trajtenberg M., Fogarty M.S. (2000) "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors", *American Economic Review*, vol 90, no. 2,pp. 215-218

Katz J.S., Martin B.R. (1997), "What is research collaboration?", *Research Policy* 26, pp. 1-18

Leopold E., May M., Paaß (2004); Data Mining and Text Mining for Science & Technology Research. In Moed H.F., Glänzel W., Schmoch U., Handbook of Quantitative Science and Technology Research. Kluwer. Dordrecht, Ch. 8.

McSherry C. (2003), *Who Owns Academic Work*, Harvard University Press

Merton R.K. (1988), "The Matthew Effect in Science, II. Cumulative Advantage and the Symbolism of Intellectual Property", *ISIS* 79, pp. 606-623

Meyer M., Bhattacharya S. (2004), "Commonalities and differences between scholarly and technical collaboration. An exploration of co-invention and co-authorship analyses", *Scientometrics* 61, pp. 443-456

Meyer M., Sinilainen T., Utecht J.T. (2003), "Toward hybrid Triple Helix Indicators: A study of university-related patents and a survey of academic inventors", *Scientometrics*, vol 58, pg 321-350

Mowatt G. et al. (2002), "Prevalence of Honorary and Ghost Authorship in Cochrane Reviews", *Journal of the American Medical Association* 287, pp. pp.2769-2771

Mowery D., Nelson R.R., Sampat B. N., Ziedonis A. (2004), *Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole Act in the United States*, Stanford University Press

Murray F. (2004), "The role of academic inventors in entrepreneurial firms: sharing the laboratory life", *Research Policy* 33, pp.643-659.

Murray F., Stern S. (2004), "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? Evidence from Patent-Paper Pairs", paper presented at the NBER Summer Institute: *Academic Science and Entrepreneurship: Dual Engines of Growth?,* 23 July, Cambridge MA

OECD (2003), *Turning Science into Business: Patenting and Licensing at Public Research Organisations*, Organization for Economic Co-operation and Development, Paris

PVA-MV (2003), *Report on the abolition of the German professors privilege: Overview of changes and challenges*, Vinnova - Swedish Agency for Innovation Systems, Stockholm

Stephan P. (1996), "The Economics of Science," *Journal of Economic Literature*, Vol XXXIV pp. 1199-1235.

Vinarov S.D. (2003), "Patent protection for structural genomics-related inventions", *Journal of Structural and Functional Genomics* 4, pp. 191-209

Weeks W.B., Wallace A.E., Kimberly B.C.S. (2004), "Changes in authorship patterns in prestigious US medical journals", *Social Science & Medicine* 59, pp.1949-1954

Tables and Figures.

Table 1

Number of academic inventors by field and years of birth

	Years			
Fields	1925-1939	1940-1954	1955-1969	Total
C	15	24	21	60
E	12	30	17	59
C	10	21	14	45
K	4	24	26	54
Total	41	99	78	218

Fields: C: Pharmacology, E: Biology, I: Chemical eng. & Materials Technology, K: Electronics and Telecom

Table 2.

Number of patents by priority date and type of assignee

	TYPE				
Years	INDIVIDUAL	OPEN	PRIVATE	n.a.	Total
1988-1991	37	29	115	22	203
1992-1994	3	33	108	25	169
1995-1997	12	41	144	22	219
1998-2000	5	10	62	9	86
Total	57	113	429	78	677

The total number of patents considered is 389. Here there is double counting because there can be many applicants for each patent.

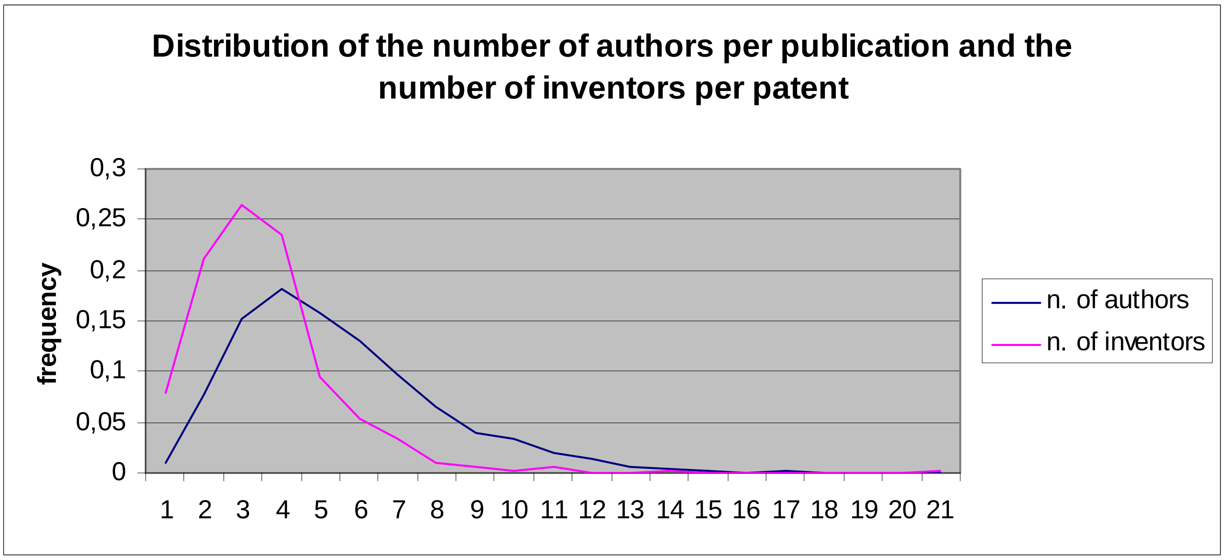
Table 3.

Number of publications by year of publication and fields

	Fields				
Years	C	E	I	K	Total
1990-1993	195	263	121	144	723
1994-1996	282	407	169	278	1136
1997-1999	229	340	87	301	957
2000-2002	43	37	16	92	188
Total	749	1047	393	815	3004

Fields: C: Pharmacology, E: Biology, I: Chemical eng. & Materials Technology, K: Electronics and Telecom. The total number of publications considered is 2838. Here there is double counting because there can be many authors for each publication.

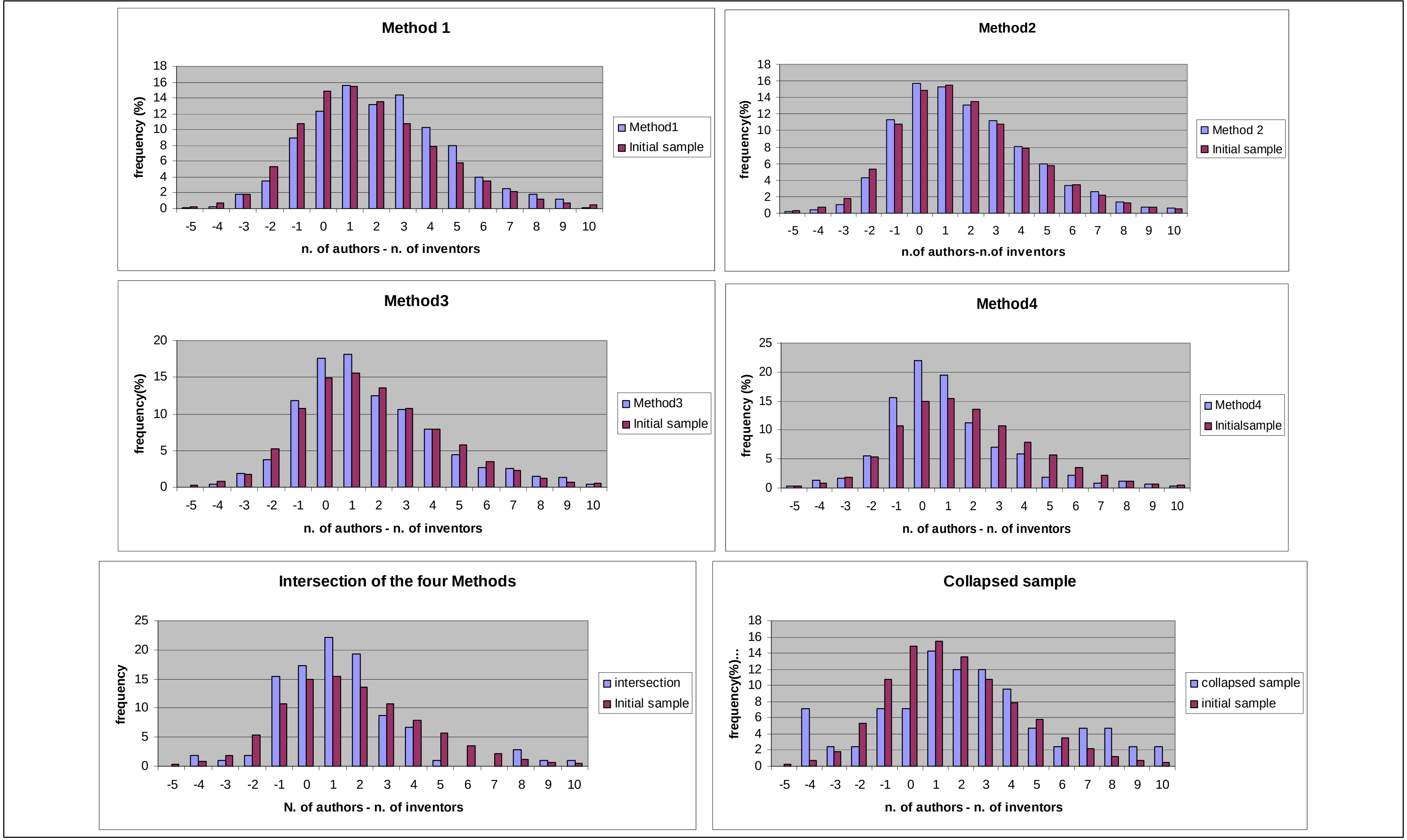
Figure 1: Distribution of number of authors per publication and number of inventors per patent.



Note: This figure refers to the total sample of 2838 publications and 389 patents. The maximum number of co-inventors is 21. There are 23 publications with a number of authors greater than 21 that are not included in the figure.

Panel 1.

The observed frequency of the gap between the number of authors and the number of inventors with the four methods of selection of the patent-publication pair.



Panel 2.

The observed frequency of the gap between the number of authors and the number of inventors in the four technological classes using Method 4 of selection of the patent-publication pair.

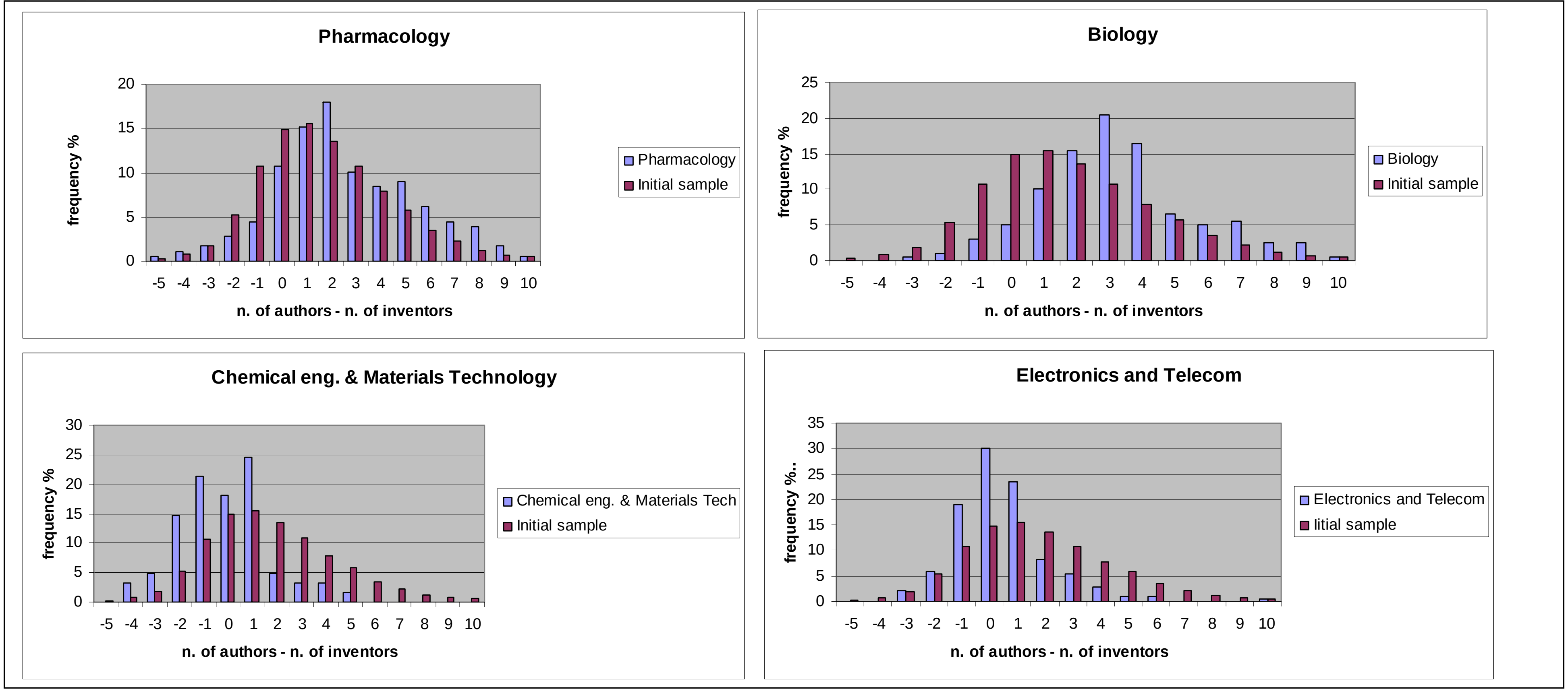


Table 4:

Summary statistics of authors and inventors by method of selection of the patent-publication pairs and by disciplines

	Average n. of authors	MIN n. of authors	MAX n. of authors	Average n. of inventors	MIN n. of inventors	MAX n. of inventors	Average difference between authors and inventors(a)	MIN difference between authors and inventors(a)	MAX difference between authors and inventors(a)	Median difference between authors and inventors(a)
All potential Pairs	8,51	1	517	3,62	1	21	4,89	-18	515	1
M1	5,43	1	44	3,34	1	21	0,40	-18	40	2
M2	10,57	1	517	3,65	1	21	6,92	-14	515	2
M3	5,00	1	42	3,36	1	21	1,64	-18	37	1
M4	5,31	1	82	3,50	1	21	1,81	-18	80	1
M5	4,40	2	13	3,11	1	11	1,30	-4	10	1
Pharmacology (M3)	6,47	2	14	3,75	1	10	2,71	-5	12	2
Biology (M3)	6,32	2	42	3,60	1	21	2,72	-18	37	3
Chemical eng. & Materials Technology (M3)	4,54	1	8	4,67	2	11	-0,13	-4	5	0
Electronics and Telecom (M3)	3,63	1	19	2,99	1	6	0,63	-3	16	0

(a) This columns refers to average min max and median value of the difference between the number of authors and inventors across each patent-publication pair.

Table 5.

Count of exclusions by position in the by-line of the excluded author and number of authors per publication. Method 1.

METHOD1		Position of the excluded author in the by-line																
Number of authors	Number of publications	1	2	3	4	5	6	7	8	9	10	11	12	13	14	All	None	Sum of exclusions
2	23	10	9													1	3	19
3	67	38	43	27												1	3	108
4	99	60	66	60	36											4	5	222
5	76	53	60	53	47	28										4		241
6	77	43	58	58	54	43	37									4	1	293
7	82	53	63	56	58	54	54	46								10	1	384
8	27	18	18	21	21	22	16	12	15							4		143
9	49	24	40	38	37	40	39	35	36	15						4		304
10	9	5	6	6	6	6	6	6	5	4	7					2		57
11	12	10	6	12	10	11	11	11	12	10	10	5						108
12	12	10	11	11	11	12	12	12	9	7	9	8	6					118
13	2	1	1	2	1	2	2	1	2	2	2	1	0	1				18
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1			13

Table 6

Count of exclusions by position in the by-line of the excluded author and number of authors per publication. Method 2.

METHOD2		position of the excluded author in the by-line																
Number of authors	Number of publications	1	2	3	4	5	6	7	8	9	10	11	12	13	14	All	None	Sum of exclusions
2	113	49	52														12	101
3	395	259	286	127												6	13	672
4	449	317	327	328	159											4	11	1131
5	405	309	310	295	284	204										11		1402
6	305	184	219	222	232	231	182									12	3	1270
7	242	157	190	179	181	186	173	179								13	2	1245
8	161	108	121	133	136	139	114	108	118							4		977
9	140	103	111	115	120	118	106	99	104	84						5		960
10	123	76	100	78	95	95	94	98	84	88	86					16		894
11	83	56	69	60	71	70	70	73	73	73	63	49				6		727
12	55	42	46	45	47	47	46	48	46	40	41	42	45			3		535
13	25	12	22	23	17	21	20	21	20	22	22	18	18	19		2		255
14	3	1	1	1	2	2	1	2	2	2	1	2	2	2	2	1		23

Table 7

Count of exclusions by position of the excluded author in the by-line and number of authors per publication. Method 3.

METHOD3		Position of the excluded author in the by-line																
Number of authors	Number of publications	1	2	3	4	5	6	7	8	9	10	11	12	13	14	All	None	sum of exclusions
2	19	9	6													1	3	15
3	97	47	63	35												1	6	145
4	98	52	57	68	39											3	6	216
5	69	40	49	50	44	29										3		212
6	58	29	37	40	44	37	26									1	3	213
7	51	25	37	27	38	36	36	26								6	1	225
8	19	13	12	13	15	14	12	9	9							2		97
9	26	12	21	19	18	20	20	20	17	8						1		155
10	8	4	4	6	6	6	5	6	5	3	4					2		49
11	13	10	9	12	11	12	12	12	13	11	11	5						118
12	14	7	9	6	10	11	13	13	12	11	10	11	11			1		124
13	2	0	0	1	0	1	1	0	1	1	1	0	0	0		1		6
14	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1			10

Table 8

Count of exclusions by position of the excluded author in the by-line and number of authors per publication. Method 4.

METHOD 4		position of the excluded author in the by-line																
Number of authors	Number of publications	1	2	3	4	5	6	7	8	9	10	11	12	13	14	All	None	Sum of exclusions
2	12	5	7															12
3	107	67	77	31												1	5	175
4	96	58	59	60	46											5	4	223
5	106	79	82	85	69	37										3		352
6	53	43	37	40	39	37	26									2		222
7	34	13	18	17	23	20	19	18								7	1	128
8	7	3	4	5	5	5	5	3	3							2		33
9	8	7	7	4	6	4	5	4	3	4						1		44
10	5	2	3	4	4	4	3	4	4	2	4					1		34
11	6	5	4	6	5	5	6	5	6	6	6	3						57
12	6	3	5	3	5	6	6	6	4	5	4	5	5					57
13	3	1	1	1	1	2	2	1	2	2	2	1	1	1		1		18
14	2	1	2	1	2	2	2	2	1	2	2	2	2	1	2			24

Table 9
Count of exclusions by position of the excluded author in the by-line and number of authors per publication. Method 5.

METHOD 5		position of the excluded author in the by-line																
Number of authors	Number of publications	1	2	3	4	5	6	7	8	9	10	11	12	13	14	All	None	Sum of exclusions
2	1	0	1															1
3	17	6	11	8													2	25
4	14	3	7	10	7												2	27
5	12	6	8	8	7	4												33
6	7	4	3	4	5	6	2											24
7	5	2	2	1	4	3	3	3									1	18
8	1	1	0	1	1	1	1	1	0									6
9	2	2	2	0	2	1	1	1	0	0								9
10	1	0	0	0	0	0	0	0	0	0	0					1		0
11	1	1	1	1	1	0	1	1	1	1	1	1						10
12	3	2	2	2	3	3	3	3	2	2	2	3	2					29
13	1	0	0	1	0	1	1	0	1	1	1	0	0	0				6
14	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-

APPENDIX: Examples of patent-publication pairs selected with M1, M2, M3 and M4

Method 1

	PATENT	*PUBLICATION*
CODE	EP457738	ISI:A1991FD66000019
TITLE	Di-aminated derivative of SV-IV protein and method for the preparation thereof	MASS-SPECTROMETRIC IDENTIFICATION OF THE AMINO DONOR AND ACCEPTOR SITES IN A TRANSGLUTAMINASE PROTEIN SUBSTRATE SECRETED FROM RAT SEMINAL-VESICLES
N. INVENTORS/AUTHORS	4	7
ABSTRACT	A DERIVATIVE OF THE SV-IV PROTEIN, WHICH HAS THE TWO ONLY GLUTAMINO RESIDUES OF SAID PROTEIN IN THE 9 AND 86 POSITIONS, EACH ONE OF SAID RESIDUES BEING COVALENTLY BONDED TO A PRIMARY AMINE, AND IN PARTICULAR TO A POLYAMINE WHICH IS PREFERABLY SPERMIDINE, AND A PROCESS FOR THE PRODUCTION OF SAID DERIVATIVES; STARTING FROM THE NATIVE SV-IV PROTEIN.	Four different transglutaminase-modified forms of a protein secreted by the rat seminal vesicles (SV-IV) were synthesized in vitro and characterized. FAB maps of both the native protein and its derivatives, produced by the purified guinea pig liver enzyme in the presence or absence of the polyamine spermidine, were obtained by mass spectrometric analysis after proteolytic digestions. Two differently derivatized SV-IV molecular forms, both possessing only one glutamine residue out of two (Gln-86) cross-linked to endogenous lysine residues, were produced when spermidine was omitted from the reaction mixture: (i) an insoluble homopolymer in which Lys-2, -4, -59, -78, -79, and -80 were involved in the linkage; (ii) a soluble form of the protein with an intramolecular epsilon-(gamma-glutamyl)lysine isopeptide bond between Gln-86 and Lys-59. Two species of SV-IV-spermidine adducts were obtained when the protein was treated with transglutaminase in the presence of high concentrations of the polyamine. The first one was characterized by one spermidine molecule covalently bound to Gln-86 and the second one by two spermidine molecules respectively bound to Gln-9 and Gln-86.

Method 2

	PATENT	*PUBLICATION*
CODE	EP1012301	ISI:000074208600018
TITLE	TOTAL SYNTHESIS AND FUNCTIONAL OVEREXPRESSION OF A $i(CANDIDA RUGOSA) LIP1 GENE CODING FOR A MAJOR INDUSTRIAL LIPASE	Design, total synthesis, and functional overexpression of the Candida rugosa lip1 gene coding for a major industrial lipase
N. INVENTORS/AUTHORS	5	5
ABSTRACT	THE DIMORPHIC YEAST CANDIDA RUGOSA HAS AN UNUSUAL CODON-USAGE WHICH HAMPERS THE FUNCTIONAL EXPRESSION OF GENES DERIVED FROM THIS YEAST IN A CONVENTIONAL HETEROLOGOUS HOST. LIPASES PRODUCED BY THIS YEAST ARE EXTENSIVELY USED IN INDUSTRIAL BIOCONVERSIONS, BUT COMMERCIAL LIPASE SAMPLES CONTAIN SEVERAL DIFFERENT ISOFORMS ENCODED BY THE LIP GENES FAMILY. IN A FIRST LABORIOUS ATTEMPT THE LIP1 GENE, ENCODING THE MAJOR ISOFORM OF THE C. RUGOSA LIPASES (CRLS), WAS SYSTEMATICALLY MODIFIED BY SITE-DIRECTED MUTAGENESIS TO GAIN FUNCTIONAL EXPRESSION IN S. CEREVISIAE. AS ALTERNATIVE APPROACH, THE GENE (1688 BP) WAS COMPLETELY SYNTHESISED WITH AN OPTIMISED NUCLEOTIDE SEQUENCE IN TERMS OF HETEROLOGOUS EXPRESSION IN YEAST AND SIMPLIFIED GENETIC MANIPULATION. THE SYNTHETIC GENE WAS FUNCTIONALLY OVEREXPRESSED IN PICHIA PASTORIS. THE RECOMBINANT CRL WAS PRODUCED AT HIGH LEVEL AND PURITY, ACCOUNTING FOR 90-95 % OF THE SECRETED PROTEINS. THE PHYSICAL-CHEMICAL AND CATALYTIC PROPERTIES OF THE RECOMBINANT LIPASE WERE COMPARED WITH THOSE OF A COMMERCIAL, NON-RECOMBINANT C. RUGOSA LIPASE PREPARATION.	The dimorphic yeast Candida rugosa has an unusual codon usage that hampers the functional expression of genes derived from this yeast in a conventional heterologous host. Commercial samples of C. rugosa lipase (CRL) are widely used in industry, but contain several different isoforms encoded by the lip gene family, among which the isoform encoded by the gene lip1 is the most prominent. In a first laborious attempt, the lip1 gene was systematically modified by site-directed mutagenesis to gain functional expression in Saccharomyces cerevisiae. As alternative approach, the gene (1647 bp) was completely synthesized with an optimized nucleotide sequence in terms of heterologous expression in yeast and simplified genetic manipulation. The synthetic gene was functionally expressed in both hosts S. cerevisiae and Pichia pastoris, and the effect of heterologous leader sequences on expression and secretion was investigated. In particular, using P. pastoris cells, the synthetic gene was functionally overexpressed, allowing for the first time to produce recombinant Lip1 of high purity at a level of 150 U/mL culture medium. The physicochemical and catalytic properties of the recombinant lipase were compared with those of a commercial, nonrecombinant C. rugosa lipase preparation containing lipase isoforms.

Method 3

	PATENT	*PUBLICATION*
CODE	EP908895	ISI:000077769000017
TITLE	Controlled hot-electron writing method for non-volatile memory cells	A new and flexible scheme for hot-electron programming of nonvolatile memory cells
N. INVENTORS/AUTHORS	3	4
ABSTRACT	IN ORDER TO OPTIMISE WRITING OF THE CELL, THE LATTER IS WRITTEN IN A CONDITION OF EQUILIBRIUM BETWEEN THE INJECTION CURRENT (IG) AND THE DISPLACEMENT CURRENT (CPPVSL). IN THIS WAY, DURING WRITING, THE VOLTAGE OF THE FLOATING GATE REGION (VFL) REMAINS CONSTANT, AS DOES THE DRAIN CURRENT AND THE RISE IN THE THRESHOLD VOLTAGE. IN PARTICULAR, BOTH FOR PROGRAMMING AND FOR SOFT-WRITING AFTER ERASURE, THE SUBSTRATE OF THE CELL IS BIASED AT A NEGATIVE VOLTAGE (VSB) WITH RESPECT TO THE SOURCE REGION, AND THE CONTROL GATE REGION OF THE CELL RECEIVES A RAMP VOLTAGE (VCG) WITH A SELECTED PREDETERMINED INCLINATION (VSL) SATISFYING AN EQUILIBRIUM CONDITION..	A new hot electron writing scheme for Flash EEPROM's is proposed that combines a positive source to bulk voltage and a ramped voltage on the control gate, The scheme exploits the equilibrium between hot electron injection and displacement current at the floating gate electrode in order to achieve a transient regime where the drain current of the cell is virtually constant. The new method allows to accurately control the threshold voltage and the programming drain current that is essentially determined by the slope of the control gate ramp and can thus be traded off with programming time over a wide range of values, The main features of the new scheme are experimentally demonstrated on up-to-date 0.6 mu m stacked gate Flash EEPROM devices.

Method 4

	PATENT	*PUBLICATION*
CODE	EP1014565	ISI:000078327700016
TITLE	Low-noise amplifier stage with matching network	A 1.5-V high drive capability CMOS op-amp
N. INVENTORS/AUTHORS	3	3
ABSTRACT	THE AMPLIFIER STAGE (50) COMPRISES A FIRST (2) AND A SECOND (3) TRANSISTOR, CONNECTED IN SERIES TO EACH OTHER BETWEEN A FIRST (4) AND A SECOND (5) REFERENCE POTENTIAL LINE. THE FIRST TRANSISTOR (2) HAS A CONTROL TERMINAL (10), CONNECTED TO AN INPUT (11) OF THE AMPLIFIER STAGE (50) THROUGH A FIRST INDUCTOR (12), A FIRST TERMINAL (15), CONNECTED TO THE SECOND REFERENCE POTENTIAL LINE (5) THROUGH A SECOND INDUCTOR (16), AND A THIRD TERMINAL (17) CONNECTED TO A FIRST TERMINAL OF THE SECOND TRANSISTOR (3). THE SECOND TRANSISTOR HAS A SECOND TERMINAL (21) FORMING AN OUTPUT OF THE AMPLIFIER STAGE (50), AND CONNECTED TO THE FIRST REFERENCE POTENTIAL LINE (4) THROUGH A LOAD RESISTOR (22). TO IMPROVE THE NOISE FIGURE, A MATCHING CAPACITOR (51) IS CONNECTED BETWEEN THE CONTROL TERMINAL (10) AND THE FIRST TERMINAL (15) OF THE FIRST TRANSISTOR (2).	A novel CMOS operational amplifier with a 1.5-V power supply is presented. It is based on a folded-mirror transconductance amplifier and a high-efficiency output stage. The amplifier achieves an open-loop gain and a gain-bandwidth product higher than 65 dB and 1 MHz, respectively. In addition, a 1-V peak-to-peak output voltage into a 500-Ohm and 50-pF output load is provided with a total harmonic distortion of -77 dB, This performance was achieved using maximum aspect ratios of 120/1.2 and 360/1.2 for the NMOS and PMOS transistors, respectively, and a quiescent current as low as 60 mu A for the driver transistors.The amplifier was implemented in a standard 1.2-mu m CMOS process,vith threshold voltages around 0.8 V. It dissipates less than 300 mu W..