KU Leuven, Faculty of Economics and Business
Dept. of Management, Strategy and Innovation (MSI)

Not like the others:
frontier scientists for high-impact inventions

Thomas Schaper, Sam Arts and Reinhilde Veugelers

MSI Discussion Paper No. 2302

Not like the others: frontier scientists for high-impact inventions[†]

Thomas Schaper[1,2], Sam Arts[3], Reinhilde Veugelers[4,5,6,7,*]

ABSTRACT

Linking scientific articles in PubMed and biomedical U.S. patents assigned to firms, we study the role of scientists active at the knowledge frontier, identified as authors on recent top articles. We expect them to be brokers between new scientific findings and inventions in industry with a high technological impact. We find that inventions made by such "frontier authors" are indeed more impactful and more likely to become breakthroughs, not only compared to those made by non-author inventors, but also compared to inventions from non-top authors and non-recent top authors. We also show that inventions with frontier science as prior art are more impactful. Frontier author patents are more likely to use frontier science as prior art in their inventions and to be first users of such frontier science. Yet, while frontier author patents have a significant impact premium on their non-frontier science prior art patents, their frontier science patents are not particularly more successful compared to frontier science patents from other types of inventors. Our results suggest that closeness to frontier science for use in their inventions is only part of the story of superior impact of frontier scientists, which seems a much broader story.

JEL classification:*O30; O31; O34*

Keywords: *Industry science links; inventor-authors; frontier science; (top) technology impact.*

Funding: This work was supported by the Fonds Wetenschappelijk Onderzoek (FWO, Research Foundation – Flanders) [11B5918N], by the KU Leuven-BOF [3H200208], and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [FOR5234/ 447967785]

[†] We thank participants in seminars and presentations at KU Leuven, TUM, CISS, WOEPSR and EPIP, and particularly Dirk Czarnitzki, Hanna Hottenrott, Stijn Kelchtermans, Fabio Montobbio, Dennis Verhoeven and Martin Watzinger for many useful comments and suggestions.

[1] Technical University of Munich, TUM School of Management, Dept. of Economics & Policy, Arcisstraße 21, D-80333 Munich, Germany. E-mail address: thomas.schaper@tum.de.
[2] KU Leuven, Faculty of Economics and Business, Dept. of Management, Strategy and Innovation, Naamsestraat 69, B-3000 Leuven, Belgium.
[3] KU Leuven, Faculty of Economics and Business Dept. of Management, Strategy and Innovation, Hendrik Conscienceplein 8, B-2000 Antwerp, Belgium. E-mail address: sam.arts@kuleuven.be.
[4] KU Leuven, Faculty of Economics and Business, Dept. of Management, Strategy and Innovation, Naamsestraat 69, B-3000 Leuven, Belgium. E-mail address: reinhilde.veugelers@kuleuven.be.
[5] Bruegel, Brussels, Belgium.
[6] PIIE, Washington, D.C., United States.
[7] CEPR, London, United Kingdom.
[*] *corresponding author.*

1 INTRODUCTION

The economics of innovation literature has since long recognized the contribution of new scientific ideas to technological progress in industry (Nelson, 1962; Griliches, 1986; Jaffe, 1989; Mansfield, 1995; Cohen, Nelson, & Walsh, 2002; Fleming & Sorenson, 2004). A vast body of empirical work has demonstrated the importance of scientific research for industrial R&D, yet also made clear that the links between science and technology are neither obvious, direct nor immediate (among others, Stephan (1996); Narin, Hamilton, & Olivastro (1997); Ahmadpoor & Jones (2017); Iaria, Schwarz, & Waldinger (2018)). When looking at channels through which science is used for technology development, inventor-authors, who cross the boundaries between scientific research and technology development in industry, have been identified as important connectors for establishing industry science links (Gittelman & Kogut, 2003; Bonaccorsi & Thoma, 2007; Breschi & Catalini, 2010).

More recently, the literature has started to dig deeper into *which* science would matter most for technology development (e.g., Poege, Harhoff, Gaessler, & Baruffaldi (2019); Veugelers & Wang (2019)), *which* individuals can effectively translate science into technology development (e.g., Arts & Veugelers (2020)), and with *which* timing such links turn out to be most viable in industry (e.g., Arora, Belenzon, & Dionisi (2021))? Of particular interest therein is to understand the importance of "frontier science" — i.e., the newest top science — for technological progress in industry, and which scientists (or engineers) would be best at linking such frontier science to successful inventions.

In this paper, we argue that inventions made by scientists who are at the scientific frontier, having produced recent top scientific articles, — whom we label as "frontier authors" — are more impactful and more likely to become "technology hits". An inherent advantage that we ascribe to inventions from frontier authors lies in their ability to identify and use the newest top scientific knowledge for their inventions.

To study the relation between inventor-authors, scientific prior art and high-impact inventions in the biomedical industry, we match all authors on scientific articles in PubMed to all inventors on U.S.

patents (cross-linking data from Torvik (2018) and Li et al. (2014)). We evaluate our research questions on the sample of all biomedical patents assigned to firms between 1980–2009, and articles in Science Citation Index (SCI)[8] listed journals. We identify *frontier science* as recently published in "top-general journals" — the most prestigious, general-interest-oriented journals in the biomedical discipline — and *frontier inventor-authors* as those inventors with recent articles in top-general journals. We quantify the contribution of inventions to technological progress by their impact on future inventions, measured by the amount of follow-on patent citations received.

We find that firm patents invented by "frontier authors" have a higher technology impact compared to similar patents without authors and, in particular, a much higher probability to yield hit patents (i.e., ranking in the upper fifth percentile of most cited patents in a technology class). Frontier author patents also have significantly higher impact compared to similar patents invented by authors missing the frontier quality — i.e., who either published in a top-general journal but not recently, or who recently published but not in a top-general journal, or who published neither recently and nor in a top-general journal. These impact premia hold stronger for first authors and decrease substantially when using lower benchmarks for assessing journal or article quality and recency of the scientific frontier. These findings are consistent with the idea that being active at the scientific frontier matters for fostering technological progress.

Digging deeper into the sources of the impact premium of frontier author patents, we further consider the type of science used in firm patents. Building on the findings of Poege, Harhoff, Gaessler, & Baruffaldi (2019), who show that the quality of the science used as prior art positively relates to the value of inventions, we examine whether frontier author patents use different science — as measured by scientific non-patent references (SNPRs) — and whether this contributes to their technology impact premium. We find that patents invented by frontier authors are not only more likely to cite scientific prior art, in general, and more extensively do so, but also that they are particularly more likely to cite frontier science, compared

[8] The Science Citation Index, alternatively denoted as Science Citation Index Expanded, collects the about 9,000 most relevant scientific journals across 178 fields. It is accessible through Clarivate Web of Science.

to patents invented by non-authors and by authors who are not at the frontier. We further show that reliance on frontier science positively relates to more impactful inventions and a greater likelihood of technology hits. But as this holds in general — i.e., not only for frontier author patents — and as, conversely, frontier author patents also outperform among those without frontier science references, their more likely use of frontier science can only be part of the story of why frontier scientists' inventions have superior impact. Finally, while frontier author patents are fast in relying on new top science, such early links to top science do not matter for technology impact — and we do not find any evidence for frontier author patent premia to depend on their priority (i.e., being first to cite frontier science), but it does matter for private value.

We document the robustness of our results across a range of alternative approaches and measures in online supplementary materials, including impact indicators other than patent citations, reflecting the private value of patents, and using semantic analysis as alternative measurement to SNPRs for frontier science links.

In their entirety, our results illustrate the key role of frontier scientists as brokers between new scientific findings and technology applications in industry. Frontier scientists foster technological progress in industry, and this effect is at least partially driven by their ability to identify and build on new scientific insights at the frontier of knowledge. Yet, our results indicate that higher impact also persists for those patents of frontier authors that do not refer to frontier science as prior art, suggesting that differences in use of science only partially captures the contribution of frontier scientists to superior inventive impact. Overall, our findings support the importance of unpacking the frontier nature of science for identifying the role of inventor-authors and prior art references to science in predicting more impactful technological inventions.

2. A REVIEW OF THE LITERATURE AND OUR HYPOTHESES

2. 1. Science as an input for technology

Scientific advance and technological progress are long-known to be closely coupled. Using forward citations as an indicator for technological impact, Fleming & Sorenson (2004), among others, show that patents that rely on scientific prior art have significantly more impact on follow-on invention. Scientific inputs have also been found to yield inventions with superior private returns (Griliches, 1979; Mansfield, 1991; Krieger, Schnitzer, & Watzinger, 2021).

A growing empirical literature has started to investigate which type of science is most relevant for technology. Most of these analyses focus on assessing the relevance of high quality science, measured as by being highly cited or published in top cited journals. In their study of patent references to scientific articles, Hicks, Breitzman, Hamilton, & Narin (2000) find that the top percentile of most cited papers also exhibit a much higher likelihood of receiving patent citations. Exploring the shock of WWI inhibiting access for Allied and Central scientists to new scientific knowledge from the opposite camp, Iaria, Schwarz & Waldinger (2018) find that researchers with higher pre-war citation reliance on top science from abroad suffered a strong relative decline in scientific output with technology relevance. Poege, Harhoff, Gaessler, & Baruffaldi (2019) show that non-patent references to a small share of high-quality scientific articles disproportionately account for higher technological value of patents.

2. 2. Inventor-authors as knowledge brokers between science and technological progress in industry

Starting from Arrow (1962), innovation scholars have highlighted the existence of widely disconnected communities of scientists and inventors and have investigated the mechanisms by which science can be linked to technological inventions. Prior work has stressed the key role of knowledge brokers between “small world” communities, i.e., individuals embedded simultaneously in the science and technology realms, such as “gatekeepers” (Allen, 1977) or “inventor-authors” (Balconi, Breschi, & Lissoni, 2004; Breschi & Catalini, 2010). The idea of knowledge brokers is also central in the literature focusing on strategies of firms to connect with academic scientists as a mechanism to benefit from scientific progress for corporate R&D (e.g., Cockburn & Henderson (1998); Fabrizio (2009)). In their analysis of the patents of 116 U.S. biotechnology firms, Gittelman & Kogut (2003) find that the number of follow-up citations of these patents increases in the share of scientific authors among inventors on the patent. Similar findings are reported for the field of nanotechnology by Bonaccorsi & Thoma (2007). More recently, Arts & Veugelers (2020) show that Ph.D. graduates working in industry who engage in academic boundary spanning — as proxied by scientific articles co-authored with academic scientists — invent more novel and impactful patents.

A significant gap in the literature remains regarding our understanding of *which* inventor-authors are best in brokering between science and technological progress in industry. Although excellence in the understanding and discovery of fundamental science does not necessarily translate into excellence in developing technological applications, which may require different minds and skills (e.g., Allen (1977)), the literature on firms' collaborations with scientists has focused on the importance of top scientists as knowledge brokers. As observed by several studies investigating the birth of the U.S. biotechnology industry (e.g., Zucker & Darby (1996); Zucker, Darby, & Brewer (1998)), outcomes of firms collaborating with scientists are highly skewed towards "star" scientists. Collaboration with star scientists on matters of basic research is often described as an effective indirect mechanism to absorb and integrate these ideas into technological applications (Zucker, Darby, & Armstrong, 2002; Colen, Belderbos, Leten, & Kelchtermans, 2022).

2. 3. Our hypotheses: ***Frontier*** authors and their use of ***frontier*** science for impactful inventions

Both scientific and technological progress are often depicted as advancing by a handful of big, new ideas or breakthroughs (Schumpeter, 1942; Kuhn T., 1962; Dosi, 1982; Scherer & Harhoff, 2000; Wang, Song, & Barabási, 2013). This immediately prompts the question of whether new, big ideas at the *frontier* of science are also more relevant for technology progress and particularly for technology breakthroughs? Although not explicitly examined, prior research suggests the importance of not only the high-quality but also the recency dimension of scientific ideas as most relevant prior art for technology development in industry. The positive link between recency of scientific involvement and technology outcomes has been highlighted by prior literature (e.g., Fabrizio (2009) and Gruber, Harhoff, & Hoisl (2013)). More recently, Poege, Harhoff, Gaessler, & Baruffaldi (2019) show that the technology impact of high-quality scientific articles is strongest for shorter time lags between the publication date of the cited articles and the patent filing.

Recent high quality research may not only be pushing the science frontier, but may also provide opportunities for new types of inventions, stemming from the translation of a new understanding of fundamental principles into impactful technological solutions, and particularly technological breakthroughs. Yet, for firms to identify and build on frontier science to create high-impact inventions remains challenging. Recent scientific insights at the knowledge frontier are especially difficult to absorb for technology inventions, due to the newness and implicitness of the embodied ideas and methods. This calls for an even more explicit role of knowledge brokers, inventor-authors with roots in both communities, for being able to use frontier science for technology development.

As we expect the scope for technology impact, and especially big impact, to come from the "frontier" of science — i.e., the pairing of excellence *and* recency of the scientific insights —, we single out "frontier authors-inventors" among the typically examined "star" scientists. The reason why we opt for this differentiation among star scientists is that we expect the advantage that involvement in top science provides for authors in technology development to decay with time. Indeed, as the frontier of knowledge is rapidly moving (e.g., Mukherjee, Romero, Jones, & Uzzi (2017)), following the obsolescence of their insights, also the impact advantage of star authors in technology development is likely to be discounted as time since top scientific engagement passes.

Conversely, we expect that inventions of frontier authors have a higher technology impact and are — in particular — more likely to be hits, not only compared to inventions from non-authors, but also compared to inventions from other authors, including star authors without (recent) frontier science contributions. Accordingly, we formalize our first hypothesis:

H1: *Inventions of frontier authors have a higher technology impact and are — in particular — more likely to be technology hits, compared to inventions of non-frontier authors or non-authors.*

We further explore the mechanisms underlying any inventive performance premium for frontier authors. We hypothesize that the specific advantage of frontier authors likely relates to their enhanced access to the rapidly moving frontier of scientific knowledge and the use of this frontier knowledge as prior art for their inventions, We argue that frontier scientists are best placed to absorb new ideas and methods at the scientific frontier for their inventions, their own created frontier knowledge, but not exclusively.

Frontier authors may not only have better and earlier access to frontier science. They may also possess a unique combination of skills, which allows them not only to produce frontier science, but also to identify and realize the use of new scientific insights for technology applications in industry. Inventor teams with frontier authors may, thus, be better able to use science, and particularly frontier science as prior art to create more impactful technologies and breakthroughs. We expect frontier authors to establish links to frontier science rapidly, often already before the new scientific knowledge becomes (publicly) diffused and, by this, enjoy significant first-mover advantages. Several empirical studies demonstrate that first-mover advantages may matter in building on new scientific ideas for technology development in industry. Arora, Belenzon, & Dionisi (2021) show that first-mover advantages in making references to scientific articles are related to a higher private value of firm patents, especially for scientific articles with greater utility for industrial applications.

In line with these arguments, we formulate the following hypotheses:

H2*: Inventions of frontier authors are more likely and faster to use frontier science as prior art compared to inventions of no-authors or non-frontier authors.*

H3: *Inventions of frontier authors that use frontier science as prior art have higher technology impact and are more likely to be technology hits compared to similar inventions from no-authors or non-frontier authors.*

3 RESEARCH SETTING, DATA & METHODS

3. 1. Data sources

We analyze the relationship between frontier authors, frontier science as prior art and technology impact for all biomedical utility patents assigned to firms and granted by the U.S. Patent and Trademark Office (USPTO) from 1980 to 2009.[9] The biomedical industry is a relevant context given the strong relation between science and technology (e.g., Gambardella (1992); Cockburn & Henderson (1998); Henderson, Pisano, & Orsenigo (1999); Murray (2002); Li, Azoulay, & Sampat (2017)). We define biomedical inventions based on three-digit patent classes assigned to the 'Drugs&Medical' category in the classification of Hall et al. (2001). Our main source of patent data is PATSTAT Global 2022 (autumn edition), which we combine with patent text and assignee information from PatentsView. We further collect monetary patent value estimates from Kogan, Papanikolaou, Seru, & Stoffman (2017). We restrict the sample to patents exclusively assigned to companies, excluding collaborative patents between industry and universities or public research institutes.[10] We link this sample to the USPTO disambiguated inventor database (Li, et al., 2014). Our final sample includes 237,345 patents invented by 159,994 unique inventors.

Using data from Torvik (2018), we match all inventors in our sample to authors from the *Author-ity* 2009 data (Smalheiser & Torvik, 2009; Torvik & Smalheiser, 2009), covering all authors on scientific articles indexed in PubMed.[11] We restrict the sample to articles published in journals in the Science Citation

[9] The delineation by grant years implies some truncation for the very last application years in our sample. This does not affect our estimates: All results are robust against excluding the last 3, 5, or 10 application years.
[10] Industry-university/public co-patents account for < 1% of all patents assigned to corporations.
[11] PubMed is a comprehensive online bibliographic index database for all, but not exclusively, biomedical literature. In the Supplementary Materials, we provide a discussion of coverage and generalizability of PubMed for links to science by patents in our sample.

Index (SCI) and link them to scientific field categories from Clarivate Web of Science.[12] 35% of the 159,994 unique inventors on biomedical firm patents in our sample have authored at least one scientific article.[13]

To identify the use of scientific prior art in patents, we retrieve scientific non-patent references (SNPRs) to all articles indexed in PubMed/ SCI, using data from the *PatCi* disambiguation (Agarwal, Lincoln, Cai, & Torvik, 2014) and Marx & Fuegi (2020).[14] We further extract the titles and abstract of all PubMed/ SCI articles and search the description of inventions in the text body all U.S. patents for semantic overlap with these. Finally, we source article-to-article citations from Microsoft Academic Graph (MAG, Sinha et al. (2015)), in order to determine the cumulative scientific impact of all articles in our data. In the Supplementary Materials, we provide further details on data processing.

3. 2. Measures

Patent impact and hit patents. We utilize forward citations — i.e., references received from subsequent patents — to quantify the technological impact of patents in our sample. Forward citations are a commonly used and established indicator in the literature to assess the quality of patents (Trajtenberg, 1990; Harhoff, Narin, Scherer, & Vopel, 1999). Patent citations capture technological value, given that they reflect how useful an invention is as input for subsequent patents. For each patent, we count the total number of citing patents within the USPTO over a ten-year moving window since filing of its worldwide first application (priority date). For robustness, we consider alternative versions of this indicator, based on (DOCDB) family-to-family citations, as well as using alternative time windows. In line with prior literature (e.g., Trajtenberg (1990); Ahmadpoor & Jones (2017)), we identify "hit patents" as those inventions entering the far-right tail of the measure, which we define as the upper fifth percentile in the citation distribution of their primary patent class-year cohort.

[12] These cover 58% of all publications indexed in PubMed for the same period. We use the 2017 version of the SCI.
[13] We provide more details and a quality assessment of the inventor-author link in the Supplementary Materials.
[14] References to SCI indexed articles account for 87.6% of all SNPRs to PubMed. For more details on SNPR data, see the Supplementary Materials.

For sensitivity evaluation, we extend the scope of our analyses to other dimensions of patent value beyond technological impact, namely private value of patents. First, we use data from Kogan, Papanikolaou, Seru, & Stoffman (2017) for monetary estimates of patent values based on abnormal stock market returns following the announcement of patent grants.[15] Second, we use a measure of patent scope based on the length of a patent's first independent claim, which proxies for the strength of legal protection (Kuhn & Thompson, 2019; Marco, Sarnoff, & Charles, 2019) and breadth of the inventive contribution (Akcigit & Ates, 2023, forthcoming) of a patent. Third, we use patent renewal fees payments (Pakes (1986); Schankerman & Pakes (1986); Hall & Harhoff (2012)), capturing the lower bounds of patent values through patent holders' revealed preferences. Results for these alternative measures are reported in the Supplementary Materials.

Frontier authors. To identify authors at the frontier of science in our patent sample, we leverage the rich heterogeneity in our data regarding inventors' author footprints. We define frontier science on two dimensions: 1) *top quality*, and 2) *recency*. We consider scientific articles to meet the highest quality criterion if they are published in one of the eight most prestigious journals of the biomedical discipline: The New England Journal of Medicine, The Lancet, JAMA Journal of the American Medical Association, Nature, Science, Cell, Nature Medicine, and the BMJ British Medical Journal.[16] These journals — which we denote as "top general journals" — span a broad range of topics and require submissions to surpass the highest peer review standards before acceptance for publication. Articles in such journals account for a large share of the most ground-breaking scientific discoveries across fields, with a high supra-disciplinary

[15] These are available only for a subset of our sample (36%), given that the data from Kogan, Papanikolaou, Seru, & Stoffman (2017) only cover patents by U.S.-listed public firms included in the CRSP database.
[16] Formally, we select these based on: 1) Being in the top category (Q1) of journals in a biomedical Web of Science field category according to the Clarivate Journal Citation Report (2018), 2) covering a broad range of fields and topics — i.e. being general-interest — and 3) ranking in the ninetieth percentile of average journal impact factor (JIF) among journals in the overall top category of biomedical journals, across fields. We exclude review journals from this ranking.

relevance.[17] We consider scientific articles to be recent if they were published within the last three years prior to the application of a focal patent.[18]

Subsequently, we classify patents as "frontier author patents" if at least one of their inventors authored a top-general journal article within the three years preceding the first patent application. We also identify patents who list non-frontier authors (either non-top science authors, or non-recent top-science authors). We examine various robustness scenarios for measuring frontier authorship, with results reported in the Supplementary Materials.

Frontier SNPRs. We trace the use of frontier science in inventions by exploring the scientific prior art references made by patents in the sample.[19] Although not a perfect measure, SNPRs are a good proxy for a genuine and important relation between the cited scientific idea as prior art input for the development of the citing invention (Tijssen, 2001; Roach & Cohen, 2013; Callaert, Pellens, & Van Looy, 2014), and are commonly used in the literature.

We rely on the patents-to-PubMed link to differentiate among types of SNPRs. Similar to the definition of frontier authors, we classify patents as "frontier SNPR patents" if at least one of their scientific references links to an article in a top-general journal which has been published in the three years leading up to the first patent application. Again, we also account for non-frontier SNPRs. For each patent–article pair within the categories of frontier and non-frontier SNPRs we further determine binary indicators for subcategories regarding timing (priority) and closeness of inventors to referenced articles. Specifically, we define patents as "first to cite" if they are the first among USPTO patents to cite a given article, in line with Arora, Belenzon & Dionisi (2021).

[17] The use of journal based indicators of scientific performance, rather than citations, has the advantage of being unaffected by future success of patents possibly affecting the top science status of inventor-authors.
[18] The three-year cut-off is a relatively strict definition of recency of scientific involvement, given that the patent development process — the actual inventive search — usually precedes the filing date for a patent by 1–2 years.
[19] We exclude SNPRs added by patent examiners. This information is, however, only available for the subset of patents granted after Jan 2001. Including examiner-given references does not change our results.

As an alternative method to relying on SNPRs for detecting frontier science links in patents, we study semantic references to new scientific concepts in the summary description of inventions in the text body of patents, with results reported in the Supplementary Materials. Similar to Iaria, Schwarz, & Waldinger (2018), we proxy "frontier scientific terms" by non-common language terms appearing for the first time in titles and abstracts of scientific articles in recent top-general journals. Subsequently, we search for reuses of these "frontier scientific terms" in the text of all 4.1 million U.S. patents granted until the end of our sample period.[20] For more information on our text-based approach, see the Supplementary Materials.

3. 3. Sample and summary statistics

Almost 60% of patents in our sample have at least one inventor with prior scientific involvement, confirming the strong connect between science and technology in biomedical research.[21] About 17% of patents are invented by authors with prior top-general journal articles, but only 7% rank in the category of frontier author patents — i.e. having at least one inventor with *recent* top-general journal performance.[22] [23]

Table 1 summarizes our main measures by type of inventor author-status of patents in the sample.[24] Patents with frontier authors receive fewer forward citations compared to patents with non-frontier authors, but also — and in particular — to patents with non-authors. Yet, they are more likely to become hit patents and are associated with higher average monetary value compared to patents in the other two groups (all pairwise differences significant below the 1%-level). Frontier-author patents are more likely and faster to

[20] For this, we consider all 11.5 million articles in PubMed/ SCI between 1878 and 2009, and search for new words — excluding signs, numbers and frequently used terms in the English language — in articles published during the period of 1970 to 2009. We, furthermore, rank scientific concepts underlying new words by quality based on the (highest) journal category in which they appear in the year of their introduction. Our results are robust to using only new words appearing in titles, as in Iaria, Schwarz, & Waldinger (2018). To establish the frontier dimension, we consider as *frontier science words* : new words in top-general journals for three years since their first publication, and as *frontier science words patents* those patents reusing at least one such term within this time frame.

[21] Considering all prior publications in PubMed, rather than only those indexed in SCI journals, does not affect the share of author-patents significantly. In this case, 62% of patents show an inventor-author link.

[22] For an alternative frontier science definition based on the top 5% most cited articles in a field (irrespective of their journal), patent shares are somewhat larger (12% frontier author patents).

[23] In the Supplementary Materials we briefly discuss the selection of authors into invention and show that frontier authors tend to be more likely than other authors to become inventors (see Table SM1; Figures SM1 and SM2).

[24] We provide more detailed summary statistics and author-level characteristics in the Supplementary Materials (see Table SM2 and Table SM3).

cite scientific prior art, with shortest lags often relating to scientific articles not yet published at the time of first patent filing. They also cite more scientific prior art, and more recent scientific prior art, especially frontier science. They are also much more likely to reuse new frontier science words in the description of their inventions. When splitting up patents by types of SNPRs, a similar pattern as for authors emerges: patents with frontier SNPRs receive fewer absolute citations, but are more likely to be top of their class compared to patents with non-frontier SNPR patents and patents without SNPRs.

Table 1: Summary statistics

VARIABLES	All	Frontier author patents	Non-frontier author patents	No-author patents
Patent forward citations (10y)	17.28	14.15	16.39	18.93
Hit patent	0.07	0.08	0.07	0.06
$-value (in mio USD)	20.52	24.80	22.38	16.73
Scientific reference (SNPR)	0.40	0.78	0.50	0.21
Number of SNPRs	4.81	17.68	5.69	1.52
Frontier SNPR (top-general journal, last 3y)	0.09	0.40	0.10	0.03
Citation lag SNPR (years)	2.08	-0.31	2.00	3.86
Citation lag to frontier SNPR (years)	0.37	-0.18	0.62	0.64
Reuse of frontier science word	0.05	0.14	0,05	0,03
Patent forward citations if SNPR = 1	15.60	13.70	14.58	19.89
Patent forward citations if frontier SNPR = 1	14.50	13.79	13.99	18.62
Hit patent if SNPR = 1	0.08	0.09	0.08	0.07
Hit patent if frontier SNPR = 1	0.09	0.09	0.09	0.09
Share of patents	1.00	0.07	0.52	0.41

Notes: The table reports patent-level group means. Sample are all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). “Frontier authors" are inventors that authored an article in a "top-general journal” in the past three years before the first patent application. "Frontier SNPRs" are references to articles in "top-general journals” published in the same three-year window (incl. articles published after first patent filing)."frontier science words" are new scientific words in "top-general journals”, appearing for the first time three years or less prior to the first patent application. For details on data sources and measures, see Section 3.

3. 4. Estimation methods

To assess the contribution of frontier authors to technology impact relative to non-frontier authors and non-authors (H1), we estimate the following generalized regression for the patents in the sample:

$$Y_i = \beta_1 FrontierAuthor_i + \sum_{j=1}^{n} \beta_n \, NonFrontierAuthor_{j(i)} + \boldsymbol{X}_i^{'}\gamma + \delta_{c(i)t(i)} + \eta_{s(i)d(i)} + \epsilon_i \tag{1}$$

We estimate equation (1) with a dependent variable Y_i capturing technology impact through counts of forward citations received by patent *i*, and, alternatively, as a binary indicator for the likelihood of patent *i* to become a hit patent.[25] $FrontierAuthor_i$ is a dummy variable which takes on the value 1 if a patent lists a frontier author as inventor. We further include a set of *n* exclusive category indicators accounting for non-frontier authors on patents, where $NonFrontierAuthor_{j(i)} = 1$ if patent *i* includes non-frontier author type *j* — i.e., non-recent authors in top-general journals, authors in top field-journals, authors in non-top journals — and 0 otherwise. The coefficient β_1 captures the difference in citation rates and top impact probability for patents with frontier authors relative to non-author patents. For assessing any differences between frontier and non-frontier author inventors, we compare β_1 with $\sum_j \beta_n$.

Our specifications control for a vector of patent level covariates $\boldsymbol{X^i}$ that may affect both frontier author status and patent impact. In our baseline model, we adjust for the number of inventors listed on a patent, as it stochastically affects the likelihood of author occurrence and because patents of larger teams tend to have systematically higher impact (Wuchty, Jones, & Uzzi, 2007). We include a full set of fixed effects for the number of inventors through a series of dummy variables for each integer value for inventor team size. This allows for non-parametric adjustment, without having to impose functional form assumptions on the underlying distributions.[26] In order to absorb permanent differences and time trends in citation rates and author involvement across technological fields, we further include primary three-digit class by first application year fixed effects. Finally, we add a series of dummies for assignee firm patent stock decile ranks interacted with a categorical variable capturing rank changes (growing, shrinking, steady) relative to the previous patenting year by the same firm, assuming a 15% annual stock depreciation rate

[25] In the Supplementary Materials, we estimate equation (1) using alternative outcome measures of patent value for robustness, specifically the likelihood of patent *i* to enter the upper first percentile of a patent class-year cohort (biggest hits), its (ln) private value in millions of USD, the (ln) word count of the first independent claim and the number of times renewal fees were paid.

[26] We replace team size values above the ninety-fifth percentile with their corresponding percentile ranks.

(Hall, Jaffe, & Trajtenberg, 2005). These aim to account for variation across institutional contexts and firm-level dynamics. In the robustness analysis, we also include firm fixed effects and inventor/author level characteristics.

In models for average technology impact, where the outcome measure is count distributed, we estimate equation (1) using Poisson pseudo-maximum likelihood estimation (PPML) (Silva & Tenreyro, 2011; Silva & Tenreyro, 2006; Correia, Guimarães, & Zylkin, 2020). For regressions of top-impact likelihood, we estimate equation (1) as linear probability models with high-dimensional fixed effects (Correia, 2014).[27] We use robust variance-covariance matrices to account for heteroscedastic error terms.

In order to test predictions from our second research question (H2) — that frontier author patents are more likely to use frontier science — we estimate equation (1) as linear probability models, where the dependent variable Y_i is a binary indicator equal to 1 if a patent includes at least one frontier SNPR, and 0 otherwise. The predictors of interest are the same as for the test of the main research question (H1).

Finally, in order to test predictions from research question H3, we decompose author- and SNPR-status of patents into cross-categories and estimate the following augmented regression equation:

$$Y_i = \sum_{j=1}^{n}\sum_{k=0}^{n} \beta_n \, Author_{j(i)} SNPR_{k(i)} + \sum_{k=1}^{n} \beta_n \, NoAuthor_i SNPR_{k(i)} + \\ + \boldsymbol{X}_i' \gamma + \delta_{c(i)t(i)} + \eta_{s(i)d(i)} + \epsilon_i, \tag{2}$$

As dependent variable $Y_{i,}$ we again include forward citation counts and hit patent likelihood. For predictors of interest on the right-hand side of the specification, we decompose author-SNPR links into a set of *n* exclusive binary indicators for the j^{th} author type patent making the k^{th} type of SNPRs. *J*'s index frontier authors and non-frontier authors, and *k*'s index frontier SNPRs, non-frontier SNPRs and no SNPRs

[27] Regressions of alternative outcome measures in the Supplementary Materials are equally estimated using OLS.

patent types. The reference group consists of patents without inventor-authors and no SNPRs. In the main specifications, we include fixed effects for both the number of inventors and the number of SNPRs[28], next to class by year and firm patent stocks characteristics.

4 ECONOMETRIC RESULTS

4. 1. Frontier author patents and technology impact

Table 2 presents the estimation results for the relationship between technology impact of patents and frontier author status (H1). Column 1 reports the baseline estimate for the impact differential of patents with *authors* (of any type) compared to non-author patents. The point estimate of 0.138 indicates a +14.8% difference in levels of average forward citations for patents of inventors with prior scientific authorship (significant at the 1%-level).[29] This author-patent premium is in line with findings in prior literature (Gittelman & Kogut, 2003; Bonaccorsi & Thoma, 2007). In Column 2, we test H1 by estimating differences in citation performance for patents of *frontier authors* and non-frontier authors. The point estimates reveal an important technology impact premium for patents with frontier authors, who receive + 30% more follow-up citations than patents without authors (reference group). This is significantly larger than the technology impact premium for patents with non-frontier authors, who receive on average + 14% more citations than non-author patents.[30] For the average (median) patent in our sample, this implies approximately five (two) excess citations to patented inventions of authors involved in recent top-science, compared to two (one) excess follow-up references in the case of authors not at the frontier of science at the time of invention[31].

---

[28] We replace the count of SNPRs with their corresponding percentile ranks for patents above the ninetieth percentile of SNPRs and group patents with less than two references into one dummy category.
[29] Coefficients from PPML estimation can be interpreted as semi-elasticities, identical to the log-linear model — i.e., as $exp^{(\beta)}-1$.
[30] The difference between frontier author and non-frontier author patents is significant below the 1%-level, using Wald tests for nonlinear combinations of parameter estimates.
[31] A comparison of the covariate adjusted estimates in column 2 and non-parametric means in Table 1 suggests sorting of frontier authors into inventive context where fewer patents are filed, but which are on average of higher value.

Table 2: Frontier author patents and technology impact

		PPML			OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
Author	0.138***			0.020***		
	(0.008)			(0.001)		
Frontier author (top-general journal, last 3y)		0.260***	0.270***		0.035***	0.036***
		(0.018)	(0.018)		(0.002)	(0.002)
Non-frontier author		0.129***			0.018***	
		(0.009)			(0.001)	
Author, top-general journal, > 3y			0.217***			0.027***
			(0.014)			(0.002)
Author, top field-journal, last 3y			0.190***			0.026***
			(0.013)			(0.002)
Author, top field-journal, > 3y			0.136***			0.017***
			(0.015)			(0.002)
Author, no top-journal			0.074***			0.011***
			(0.010)			(0.001)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	237,114	237,114	237,114	237,124	237,124	237,124
Pseudo/ adjusted R-squared	0.28	0.28	0.28	0.03	0.03	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of its primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

In order to more clearly tease out the frontier character of authorship, we further split up in Column 3, the category of non-frontier author patents by quality and recency of scientific backgrounds of inventor-authors. The results reveal the importance of both recency and top quality for identifying the premium for frontier science inventor-authorship: patents of frontier authors achieve considerably higher excess impact than those of authors who published in highest-ranked journals longer ago (more than three years).[32] Also the premium in impact between publishing in general top journals and in leading field-journals is substantial. Although there is still a premium for patents of authors who lack top-journal articles, this premium is considerably lower compared to patents of authors with top publications, certainly if these are frontier articles. Overall, the results support a significant technology premium for patents from top science

[32] Difference in coefficients statistically significant below the 5%-level.

authors over authors in general, but particularly from authors in recent top-general journals (frontier authors).

In Columns 4 to 6 of Table 2, we report estimations to test the second part of research question H1 — i.e. that frontier author presence is especially important for the creation of technology hits. Results imply a qualitatively analogous relationship, however effects are economically larger than for average impact: for patents with frontier authors the excess hit likelihood is about +64% higher, corresponding to a two times larger increment than estimated for patents with non-frontier authors (Column 5). The critical value of recency of scientific involvement, in particular for frontier science (+0. 9 percentage points), is reconfirmed for hit patent rates in Column 6.[33] Taken together, results in Table 2 provide strong support for H1.

We investigate robustness of these results in several ways (see Supplementary Materials). First, we estimate alternative specifications holding constant institutional context (firm fixed effects) and inventor/author level characteristics, such as the number of authors on a patent, their prior experience in patenting and publishing, as well as career tenure. Tables SM4–5 show that our baseline estimates are fully robust.

Table SM6 shows that frontier authors are even a much stronger predictor for top impact when assessing the likelihood of patents to become "biggest hits" — i.e., entering the upper 1% of patents in a class, instead of the upper 5% used in the main analysis. Table SM7 considers private values for measuring patent outcomes. Our main results are fully robust when using USD values from Kogan et al. (2017), claim length and renewal fees. Across these indicators, the performance premium for frontier author patents seems to be even more outspoken, particularly the recency dimension of frontierness, while there is consistently little to no difference in value between non-frontier and non-author patents.

Results are further robust and similar in magnitude when we define frontier authors based on recent most cited (upper 5%) articles in scientific fields rather than journal outlets (see Table SM8). We also find

[33] Difference in coefficients significant below the 1%-level.

excess patent citations — and in particular excess hit likelihood — to be significantly larger for patents of inventors listed as *first* authors on frontier science articles, consistent with the view of a rapidly moving frontier whose value for technology is magnified for authors with substantial contributions to this frontier (see Table SM9).

Table SM10 shows that estimates are qualitatively unaltered, though of larger magnitude, when considering DOCDB patent family-to-family rather than U.S. patent-to-patent forward citations. Table SM11investigates sensitivity to shorter and, respectively, longer time windows for assessing patent impact, and shows largest premia for shorter time windows, suggesting that frontier authors have, particularly, faster impact and, respectively, top impact. Tables SM12a–b show that frontier author premia of average and top impact are larger when excluding continuations. However, our results are also fully robust on the subset of only continuations (see SM12b).[34]

We also provide split-sample estimates of excess citations and hit likelihood of frontier and non-frontier author patents by technology fields (see Table SM13). In Figure SM3, we report changes of frontier author effects across different value measures over time in our sample. The frontier author patents premium is relatively stable in the long-run, however it shows more across-year variation compared to the excess rate of non-frontier author patents. Figure SM4 shows that frontier authors shifted from appearing on, on average, smaller firms' patent stocks, to patenting with larger firms' patent stocks, starting from around the year 2000. At the same time, the average inventor team size of frontier author patents appears to have significantly increased since the early 1990s, relative to non-frontier and non-author patents, all this motivating the control for firms' patent stocks and inventor team size in our analysis.

[34] We identify inventors based on the granted patent. In our sample, only continuations-in-part provoke changes to the inventors listed: In 33% of continuations-in-part an inventor is removed after priority application, in 42% of cases a new inventor is added.

4. 2. Frontier author patents and their use of frontier SNPRs

With the evidence shown in the previous section that patents of frontier authors receive considerably more follow-up citations and are especially more likely to become hit patents, in this section we further explore to which extent this impact advantage is related to a different use of scientific prior art in inventions. We particularly consider the use of frontier science as a possible mechanism related to higher technology impact for frontier author patents (H2-3).

4.2.1. Frontier SNPRs and (top) technology impact

In order to support our search for the use of frontier science as prior art as an argument for the impact advantage of frontier author patents (H3), we first document whether frontier SNPRs are indeed a potential source of higher technology impact. For this, we exchange the predictors of interest from equation (1) with dummies for types of scientific references made and evaluate their influence on the expectation of patent technology impact.[35]

[35] This is a similar approach to the estimation models by Fleming & Sorenson (2004), Poege, Harhoff, Gaessler, & Baruffaldi (2019), and Arora, Belenzon, & Dionisi (2021).

Table 3: Frontier SNPR patents and technology impact

		PPML			OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
SNPR	0.243***			0.031***		
	(0.009)			(0.001)		
Frontier SNPR (top-general journal, last 3y)		0.300***			0.035***	
		(0.024)			(0.003)	
Non-frontier SNPR		0.180***			0.019***	
		(0.015)			(0.002)	
Frontier SNPR, first			0.182***			0.022***
			(0.033)			(0.004)
Frontier SNPR, not first			0.323***			0.038***
			(0.026)			(0.004)
Non-frontier SNPR, first			0.146***			0.016***
			(0.018)			(0.002)
Non-frontier SNPR, not first			0.193***			0.021***
			(0.015)			(0.002)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of SNPRs	No	Yes	Yes	No	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	237,114	237,114	237,114	237,124	237,124	237,124
Pseudo / adjusted R-squared	0.28	0.29	0.29	0.03	0.03	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on SNPR status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 3 first confirms the literature that science-referencing patents have superior impact and higher hit patent likelihood (Columns 1 and 4). But, most importantly for our story, Table 3 shows largest impact premia for patents making references to frontier science publications, particularly for hit probabilities: patents citing a recent top-science article as prior art receive, on average, about +35% more follow-up patent citations than those without any SNPRs, an increment almost twice the size of our estimate for patents with non-frontier science references (Column 2). In addition, they are more than 60% more likely to enter the top-5% most impactful patents in a class cohort, compared to about one third of higher likelihood for non-frontier SNPR patents (Column 5). Estimates in Table SM14 show that recency appears to matter significantly more as predictor for high impact for patents with links to top science, while it only has a marginal influence for patents relying exclusively on non-top SNPRs. This pattern is in line with prior findings of Poege, Harhoff, Gaessler, & Baruffaldi (2019).

Importantly, however, Columns 3 and 6 of Table 3 show that priority — i.e., being first to cite a frontier science article — is not crucial for the technology impact of frontier science.[36] The average technology impact of patents with first-time frontier SNPRs is even significantly lower compared to those with follower frontier SNPRs. Also for hit patent probability, a penalty for priority is confirmed. This suggests, that any first-mover returns from the use of (frontier) science in technology do not affect technology impact. Table SM15 shows that these results are further fully robust when we measure frontier science links in patents by reuse of frontier science terms rather than by frontier SNPRs — i.e., patents with frontier science terms have a significantly larger technology impact — but there is no extra advantage from being first in reusing frontier science words.

In Table SM16, we compare these results to measures of private value, in line with Arora, Belenzon & Dionisi (2021). In these estimations, we also find that frontier SNPRs are strongly associated with higher private value, but — and this in contrast to the results based on patent citations — priority does translate into higher private value.

4. 2.2. Frontier author patents and likelihood of frontier SNPRs

Having confirmed in our sample that frontier SNPRs matter for impact, we proceed to examine our hypothesis that patents with frontier authors would be more likely to make frontier SNPRs (H2). Column 1 in Table 4 first shows that both frontier and non-frontier author patents are significantly more likely to reference scientific articles compared to patents by non-authors. But, most interestingly, Column 1 further shows that this is even much more likely for frontier author patents than for non-frontier author patents. Estimates from PPML count models in Column 2 imply, moreover, that this is not limited to a higher extensive margin, but that, conditional on citing science, frontier author patents also typically display a considerably larger number of SNPRs.

[36] We consider SNPRs as "first" only to articles that receive at least one subsequent U.S. patent citation in the *PatCi* or *PCS* data. Relaxing this restriction does not qualitatively alter the results.

Table 4: Frontier author patents and frontier SNPR likelihood

	OLS	PPML		OLS	
		with SNPR subsample		with author subsample	
	(1)	(2)	(3)	(4)	(5)
VARIABLES	SNPR	# SNPRs	Frontier SNPR	First Frontier SNPR	No self-cits Frontier SNPR
Frontier author	0.323***	0.868***	0.176***	0.064***	0.149***
	(0.004)	(0.020)	(0.005)	(0.008)	(0.004)
Non-frontier author	0.163***	0.316***	0.004		
	(0.002)	(0.016)	(0.003)		
Constant	0.292***	2.250***	0.197***	0.403***	0.182***
	(0.001)	(0.014)	(0.002)	(0.018)	(0.007)
With frontier SNPR subsample				Yes	No
Author-level controls				Yes	Yes
DV subsample average	0.399	12.063	0.223	0.316	0.246
Patent class x year	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes
Number of SNPRs	No	No	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes
Observations	237,124	94,223	94,223	18,330	73,908
Adjusted / pseudo R-squared	0.28	0.17	0.31	0.13	0.32

Notes: Each column reports parameter estimates from regressions of scientific reference on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures the probability that a patent includes prior art references (SNPRs) to scientific articles indexed in PubMed / SCI, in columns (1) and (3)-(5), and the number of SNPRs in column (2). Additional author-level controls in columns (4)-(5) are the logs of prior publications, patents, co-author network size, and number of authors fixed effects. Examiner-given SNPRs are excluded from patents for which this information is available (after Jan 2001). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Such disparity in volume of SNPRs implies that, ceteris paribus, frontier author patents would already stochastically be more likely to make frontier SNPRs. To evaluate, though, whether they are also more selective towards frontier science in their references, we estimate linear probabilities to make frontier SNPRs conditional on the number of SNPRs (Column 3). The coefficients in Column 3 show that patents with frontier authors are considerably more likely to cite frontier science, even when holding constant the number of articles cited — more than twice as likely compared to patents without authors. This contrasts markedly with patents of authors, which are not frontier authors. Although they are also more likely to make SNPRs (cf. Column 1), patents from non-frontier authors are not significantly more likely than non-author patents to make references to frontier science.

We test in Column 4 of Table 4, whether frontier author patents are more likely to be the first to use the frontier science being referenced. We find, indeed, that among patents from authors and with frontier SNPRs, relative to the dependent variable subsample average, those involving frontier authors are about +20% more likely to establish priority (i.e., are first to cite a given frontier science article).

Prior literature has highlighted the prevalence and weight of self-citations in inventor-author science links in technology (Tijssen, 2001; Ahmadpoor & Jones, 2017). Indeed, 11% of patents with frontier SNPRs in our sample include self-references to frontier articles, accounting for more than one in three frontier author patents with frontier SNPRs (36%). In Column 5, we discard any self-citations from the measurement of the dependent variable and reestimate differences in likelihood of frontier SNPR for the subsample of author patents. Estimates show that even when discarding self-citations, patents with frontier authors are significantly more likely to cite frontier articles (significant at the 1%-level).

Taken together, the results in Table 4 lend general support for our hypothesis that patents with frontier authors much more likely link to frontier science as prior art (H2). In the Supplementary Materials we show robustness of these results when measuring (frontier) science links in patents by the reuse of new scientific words rather than SNPRs (Table SM17).

4. 2.3. Frontier SNPRs as a source of impact-advantage of frontier author patents

In this section we report the estimates testing our hypothesis that the superior impact performance of frontier author patents can be attributed to more use of frontier science in these inventions (H3). To this end, Table 5 shows coefficient estimates for equation (2) from regressions of technology impact, this time decomposing frontier and non-frontier author patents into those making frontier, non-frontier and no SNPRs.

We first investigate any differential impact for frontier authors on SNPR patents, in general: examining the average citation impact, Column 1 in Table 5 shows that frontier authors attain significantly higher impact on inventions with SNPRs compared to their inventions without SNPRs. But this also holds for patents by non-frontier authors and non-authors. Nevertheless, the premium for frontier author patents, relative to non-frontier author and non-author patents, holds within both categories and with no significant change in magnitude— in other words: frontier author patents outdo other patents, irrespective of whether they include SNPRs or not. This also holds for hit-probabilities (Column 2), with the excess premium of SNPR patents among frontier author patents being even less pronounced.

Columns 3 and 4 further decompose SNPRs into frontier and non-frontier types. The point estimates seemingly show an impact premium for frontier SNPR patents within frontier author patents(both for average impact and hit probability), but this difference is small and non-significant. On the contrary, non-frontier authors and no-authors patents attain a significantly higher impact premium from their inventions with frontier SNPRs compared to their patents without (even more so for hit impact (Column 4)). Overall, the performance premium for frontier author patents among frontier SNPR inventions appears to be small and insignificant. When we examine patents with non-frontier SNPRs, those by frontier authors, nonetheless, do have a significant impact premium compared to non-frontier author and non-author patents. This is also true for patents without SNPRs.

Overall, the higher proclivity of frontier authors to use frontier SNPRs for their inventions helps to explain the overall higher impact of their patents, given that such frontier SNPRs are associated with generally higher impact, but is not sufficient to explain the full impact premium for frontier author patents compared to non-frontier author patents. Patents with frontier authors on board seem to outperform being able to make more often use of frontier science, but not necessarily better use of frontier science compared to non-frontier author patents and, even not, non-author patents. Yet, as they still have an edge compared to the latter regarding inventions not using frontier SNPRs, the results suggest that use of frontier SNPRs is only part of the story of superior performance of frontier authors in technology development.

When looking at private patent values, using the measure by Kogan, Papanikolaou, Seru, & Stoffman (2017) (see Table SM18), we also find that frontier author patents with frontier SNPRs have the highest excess value, higher than other frontier author patents, and also a higher premium than frontier SNPR patents from non-frontier authors and non-authors. For private patent value, these differences are generally more pronounced and statistically significant. Access to frontier science as an explanation for the impact premium of frontier author patents (H3) seems, thus, to be more confirmed for private value impact than for technology impact. Nevertheless, also for private patent value premia, access to frontier science is only part of the impact premium story, as also in these estimates, there still is a significant unexplained premium for frontier authors patents among patents without frontier SNPRs.

Table 5: Frontier author patents, frontier SNPRs and technology impact

	PPML	OLS	PPML	OLS
	(1)	(2)	(3)	(4)
VARIABLES	PatCit10	Hit patent	PatCit10	Hit patent
Frontier author, SNPR	0.339***	0.038***		
	(0.027)	(0.003)		
Frontier author, no SNPR	0.216***	0.031***	0.217***	0.031***
	(0.031)	(0.004)	(0.031)	(0.004)
Non-frontier author, SNPR	0.251***	0.030***		
	(0.017)	(0.002)		
Non-frontier author, no SNPR	0.125***	0.016***	0.125***	0.016***
	(0.010)	(0.001)	(0.010)	(0.001)
Non-author, SNPR	0.216***	0.024***		
	(0.018)	(0.002)		
Frontier author, frontier SNPR			0.401***	0.046***
			(0.034)	(0.005)
Frontier author, non-frontier SNPR			0.345***	0.039***
			(0.035)	(0.004)
Non-frontier author, frontier SNPR			0.364***	0.046***
			(0.027)	(0.004)
Non-frontier author, non-frontier SNPR			0.250***	0.030***
			(0.017)	(0.002)
No author, frontier SNPR			0.372***	0.043***
			(0.041)	(0.006)
No author, non-frontier SNPR			0.211***	0.024***
			(0.018)	(0.002)
Patent class x year	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes
Number of SNPRs	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes
Observations	237,114	237,114	237,124	237,124
Pseudo / adjusted R-squared	0.29	0.03	0.29	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on decompositions of (frontier) inventor-author status by (frontier) SNPR-type for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1) and (3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (2) and (4). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Tables SM20 and SM21, we revisit the role of priority in driving any impact premium for frontier author patents originating from links to frontier science. Although frontier author patents are more likely to be the first to use a frontier science paper (cf Table 4, Column 4), our estimates in Table SM20 (Columns 3 and 4) show that the impact-advantage of frontier author patents related to their use of frontier science does likely not stem from their ability to establish priority to frontier SNPRs (i.e., from being first to cite recent top science). Like for all other patents, there is no technology impact premium for frontier author patents from priority in using frontier science. When we consider private value (Table SM21, Column 4), frontier author patents do, however, hold a premium compared to non-frontier author patents on patents with priority to frontier science. But, then again, given that they hold a even larger premium on patents with SNPRs to frontier science *without* priority, also with regards to any private value superior performance for frontier author patents, there appears to be more to the story than what could be explained by privileged early access.

Tables SM20 (Columns 1 and 2) and SM21 (Columns 1–3) further explore the role of self-citations in driving any impact premium for frontier author patents related to the use of frontier science. Results in Table SM20 show that frontier author patents who self-cite their frontier science attain a significantly higher technology impact premium compared to other types of frontier SNPRs that frontier author patents make (although not a significantly higher hit rate). In fact, only for patents with self-cited frontier SNPRs (which are a bit more than one third of all frontier author patents with frontier SNPRs), there is a significant technology impact premium of frontier author over non-frontier author patents. Only for these patents, the prediction of H3 is confirmed — i.e., access to own frontier science seems to provide a lever for higher technology impact inventions. For other frontier science links, there is no significant technology premium for frontier author patents, refuting H3. Also when looking at private value (Table SM21), self-cited frontier science provides a significant premium over other types of frontier SNPRs for frontier author patents, but in this case — consistent with the result in Table SM18 — frontier author patents can assure a private value premium, also for non-self-cited frontier SNPRs.

5. CONCLUSIONS

In this paper, we contribute to the literature by identifying and illustrating the key role of frontier scientists who are both active at the scientific frontier — i.e., recently published in a top-general journal — and involved in technology development (as inventor on a patent) in the biomedical industry. We show that corporate inventions with frontier scientists on board have a higher technological impact on future inventions and are more likely to be breakthroughs, and this compared not only to inventions made by non-authors, but also to inventions with authors who are away from the frontier, either in time or quality. We furthermore show that this impact premium is partially driven by their ability in identifying and building on new scientific insights at the frontier of knowledge, as reflected by a higher inclination and intensity to use science as prior art for their inventions, and particularly frontier science (including, but not only, their own frontier science), the latter being a more impactful type of science for technology impact. Although frontier authors possess a first mover advantage to access frontier science, we find no evidence that this first mover advantage is linked to a higher technology impact. But our results also indicate that the technology impact premium of frontier author inventions is, evidently, a much bigger story than only access to frontier science: it holds more generally and not only among inventions where they use frontier science.

Our main results are robust to several alternative approaches, including the use of text mining techniques rather than scientific references (SNPRs) to identify the use of science in patents. When examining private value of patents, rather than technology impact through patent citations, the excess impacts of frontier author inventor patents and frontier SNPRs patents are not only confirmed, but are even more pronounced. For private value, access to frontier science— including priority access — does provide a significant premium and specifically more so for inventions of frontier authors . Yet, analogously to their technology impact premium, the superior private value created by frontier author inventions is a bigger story than only priority access to frontier science.

Our findings mainly contribute to the literature on the returns from science involvement for inventors in technology development in industry (Gittelman & Kogut, 2003; Bonaccorsi & Thoma, 2007; Gruber, Harhoff, & Hoisl, 2013). We illustrate the significant heterogeneity in this relationship by identifying the importance of the *frontier nature* of both the inventor-author link and the scientific prior art. First, we show that the recency of involvement of authors in top publications and the recency of the top scientific prior art which is used in inventions are important for identifying differences in technology impact and specifically for becoming a breakthrough. Second, we show that frontier inventor-authors are not merely "bridge-builders" between science and technology (Breschi & Catalini, 2010), but that their higher inventive performance is at, least partially, driven by their more likely identification and integration of new scientific ideas at the frontier of knowledge, including — but not only — their own frontier science. Our results, however, also indicate that there is more to their story than their access to science.

Our paper further contributes to the literature on firm participation in scientific research to increase downstream absorptive capacity (Rosenberg, 1990; Gambardella, 1992; Arora, Belenzon, & Sheer, 2021). Here, our findings establish the role of frontier authors involved in corporate technology development as a crucial mechanism for capturing value from scientific progress. Thereby, our "frontier science" perspective complements prior literature by showing recency effects of top science participation on technology search (Fabrizio, 2009). Our findings also speak specifically to recent work on channels of early access for firms to new scientific findings (Baruffaldi & Poege, 2020; Arora, Belenzon, & Dionisi, 2021). We advance this stream of work by highlighting that access to frontier science — i.e., top science when it is still recent — matters for generating superior technology impact, and that active involvement of frontier authors who are themselves actively involved in frontier science provide (first) access, but that such (first) use of frontier science is only part of the story of the superior performance impact of frontier author inventions.

Our paper could entail normative implications to the extent that our findings highlight the uneven distributions in the creation of scientific knowledge and its impact, and in particular the importance of frontier scientific research for technological progress. When the gains from these connections are usually preceded by significant risk taking and uncertainty of outcomes, this calls for particular attention to any specific barriers for frontier author inventors, in the first place, and for access to the results of frontier research for technology development, more generally. Moreover, if the involvement of corporate inventors in frontier science enables more use of frontier science in commercial technology which generate superior value, this makes a strong case for the importance of industry direct participation in frontier research, more particularly, by their willingness and ability to employ, or directly collaborate with, frontier authors. While this calls for an inclusive design of institutional regimes governing involvement of frontier scientists in corporate technology, the question for policy makers whether to intervene to ensure better institutional regimes, which channels might be most effective to target, and with which instruments, is an on-going and interesting avenue for future research.

Our results are sufficiently interesting at this stage to call for further research to dig deeper into the superior performance of frontier author and frontier science based inventions. This includes a deeper dive into who these frontier author inventors are, what makes them different from other non-authors inventors, and from other authors and star authors, which are not at the frontier. Likewise, the selection of frontier authors into teams of inventors, and into firms working with them, needs further investigation. Also the type of access and use of which type of science, as mechanism for explaining superior performance of frontier author inventions, requires a more granular analysis, mapping in greater detail the science-invention links, by using, for instance, advances in text-based methods to establish links other than SNPRs.

.

REFERENCES

Agarwal, S., Lincoln, M., Cai, H., & Torvik, V. I. (2014). Patci—a tool for identifying scientific articles cited by patents.

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science, 357*, 583–587.

Akcigit, U., & Ates, S. T. (2023, forthcoming). What happened to US business dynamism? *Journal of Political Economy*.

Allen, T. J. (1977). *Managing the flow of technology: Technology transfer and the dissemination of technological information within the R&D organization.* Cambridge, MA: The MIT Press.

Arora, A., Belenzon, S., & Dionisi, B. (2021). *First-mover Advantage and the Private Value of Public Science.* Tech. rep., National Bureau of Economic Research.

Arora, A., Belenzon, S., & Sheer, L. (2021). Knowledge spillovers and corporate investment in scientific research. *American Economic Review, 111*, 871–98.

Arrow, K. J. (1962). Economic Welfare and the Allocation of Resources. En K. J. Arrow, *The Rate and Direction of Inventive Activity: Economic and Social Factors* (págs. 609-626). New York: Universities-National Bureau Committee for Economic Research, Committee on Economic Growth of the SocialScience Research Council .

Arts, S., & Veugelers, R. (2020). Taste for science, academic boundary spanning, and inventive performance of scientists and engineers in industry. *Industrial and Corporate Change, 29*(4), 917-933.

Audretsch, D. B., & Stephan, P. E. (1996). Company-scientist locational links: The case of biotechnology. *American Economic Review, 86*(3), 641-652.

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), 127-145.

Baruffaldi, S., & Poege, F. (2020). A Firm Scientific Community: Industry Participation and Knowledge Diffusion. *Max Planck Institute for Innovation & Competition Research Paper*.

Bonaccorsi, A., & Thoma, G. (2007). Institutional complementarity and inventive performance in nano science and technology. *Research policy, 36*, 813–831.

Breschi, S., & Catalini, C. (2010). Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks. *Research Policy, 39*, 14–26.

Callaert, J., Pellens, M., & Van Looy, B. (2014). Sources of inspiration? Making sense of scientific references in patents. *Scientometrics, 98*, 1617–1629.

Cockburn, I. M., & Henderson, R. M. (1998). Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *The journal of industrial economics, 46*, 157–182.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). Links and impacts: the influence of public research on industrial R&D. *Management science, 48*, 1–23.

Colen, R. M., Belderbos, R., Leten, B., & Kelchtermans, S. (2022). Reaching for the Stars: When Does Basic Research Collaboration between Firms and Academic Star Scientists Benefit Firm Invention Performance? *Journal of Product Innovation Management*.

Correia, S. (2014). *REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects.* Boston: Statistical Software Components, Boston College Department of Economics.

Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal, 20*(1), 95-115.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy, 11*(3), 147-162.

Fabrizio, K. R. (2009). Absorptive capacity and the search for innovation. *Research Policy, 38*(2), 255-267.

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic management journal, 25*, 909–928.

Gambardella, A. (1992). Competitive advantages from in-house scientific research: The US pharmaceutical industry in the 1980s. *Research Policy, 21*(5), 391-407.

Gittelman, M., & Kogut, B. (2003). Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Science, 49*, 366–382.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, 92-116.

Griliches, Z. (1986). Productivity, R&D, and the Basic Research at the Firm Level in the 1970's. *American Economic Review, 76*, 141–154.

Gruber, M., Harhoff, D., & Hoisl, K. (2013). Knowledge recombination across technological boundaries: Scientists vs. engineers. *Management Science, 59*, 837–851.

Hall, B. H., & Harhoff, D. (2012). Recent research on the ecoomics of patents. *Annual Review of Economics, 4*(1), 541-565.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. *The NBER patent citation data file: Lessons, insights and methodological tools*. National Bureau of Economic Research Cambridge, Mass., USA.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics, 36*(1), 16-38.

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics and statistics, 81*, 511–515.

Henderson, R., Pisano, G. P., & Orsenigo, L. (1999). The pharmaceutical industry and the revolution in molecular biology: interactions among scientific, institutional, and organizational change. En R. Henderson, G. P. Pisano, L. Orsenigo, D. C. Mowery, & R. R. Nelson (Edits.), *Sources of Industrial Leadership* (págs. 267-311). New York: Cambridge University Press.

Hicks, D., Breitzman, A., Hamilton, K., & Narin, F. (2000). Research excellence and patented innovation. *Science and Public Policy, 27*, 310–320.

Iaria, A., Schwarz, C., & Waldinger, F. (2018). Frontier knowledge and scientific production: evidence from the collapse of international science. *The Quarterly Journal of Economics, 133*(12), 927–991.

Jaffe, A. B. (1989). Real effects of academic research. *The American economic review*, 957–970.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics, 132*(2), 665-712.

Krieger, J. L., Schnitzer, M., & Watzinger, M. (2021). *Standing on the shoulders of science.* Cambridge, MA: Harvard Business School Working Paper No. 21-128.

Kuhn, J. M., & Thompson, N. C. (2019). How to measure and draw causal inferences with patent scope. *International Journal of the Economics of Business, 26*(1), 5-38.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Li, D., Azoulay, P., & Sampat, B. N. (2017). The applied value of public investments in biomedical research. *Science, 356*, 78–81.

Li, G.-C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., . . . Fleming, L. (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). *Research Policy, 43*, 941–955.

Mansfield, E. (1991). Academic research and industrial innovation. *Research policy, 20*, 1–12.

Mansfield, E. (1995). Academic research underlying industrial innovations: sources, characteristics, and financing. *The review of Economics and Statistics*, 55–65.

Marco, A. C., Sarnoff, J. D., & Charles, A. W. (2019). Patent claims and patent scope. *Research Policy, 48*(9), 1037-1090.

Marx, M., & Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal, 41*(9), 1572-1594.

Mukherjee, S., Romero, D. M., Jones, B., & Uzzi, B. (2017). The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science advances, 3*, e1601315.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. *Research policy, 31*, 1389–1403.

Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between US technology and public science. *Research policy, 26*, 317–330.

Nelson, R. (1962). *The link between science and invention: The case of the transistor. The rate and direction of inventive activity.* Princeton, NJ: Princeton University Press.

Pakes, A. (1986). Patents as options: Some estimates of the value of holding European Patent stocks. *Econometrica, 54*(4), 755-784.

Poege, F., Harhoff, D., Gaessler, F., & Baruffaldi, S. (2019). Science quality and the value of inventions. *Science advances, 5*, eaay7323.

Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management Science, 59*, 504–525.

Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy, 19*, 165-174. doi:https://doi.org/10.1016/0048-7333(90)90046-9

Schankerman, M., & Pakes, A. (1986). Estimates of the value of patent rights in European countries during the post-1950. *Economic Journal, 96*(384), 1052-1076.

Scherer, F. M., & Harhoff, D. (2000). Technology policy for a world of skew-distributed outcomes. *Research Policy, 29*(4-5), 559-566.

Schumpeter, J. A. (1942). *Socialism, capitalism and democracy.* Harper and Brothers.

Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics, 88*, 641–658.

Silva, J. S., & Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters, 112*, 220–222.

Sinha, A. S., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An Overview of Microsoft Academic Service (MAS) and Applications. *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)* (págs. 243-246). New York, NY, USA: ACM.

Smalheiser, N. R., & Torvik, V. I. (2009). Author name disambiguation. *Annual review of information science and technology, 43*, 1–43.

Stephan, P. E. (1996). The economics of science. *Journal of Economic literature, 34*(3), 1199-1235.

Tijssen, R. J. (2001). Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows. *Research Policy, 30*, 35–54.

Torvik, V. I. (2018). Author-Linked data for Author-ity 2009. *Author-Linked data for Author-ity 2009*. University of Illinois at Urbana-Champaign. doi:10.13012/B2IDB-4370459_V1

Torvik, V. I., & Smalheiser, N. R. (2009). Author name disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data (TKDD), 3*, 1–29.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The RAND journal of economics*, 172-187.

Veugelers, R., & Wang, J. (2019). Scientific novelty and technological impact. *Research Policy, 48*, 1362–1372.

Wang, D., Song, C., & Barabási, A.-L. (2013). Quantifying long-term scientific impact. *Science, 342*(6154), 127-132.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036-1039.

Zucker, L. G., & Darby, M. R. (1996). Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences, 93*, 12709–12716.

Zucker, L. G., Darby, M. R., & Armstrong, J. S. (2002). Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. *Management science, 48*, 138–153.

Zucker, L., Darby, M., & Brewer, M. (1998). Intellectual Human Capital and the Birth of US Biotechnology Enterprises. *American Economic Review, 88*, 290-306.

Supplementary materials

to

Not like the others: Frontier scientists for high-impact inventions

Thomas Schaper[1,2], Sam Arts[3], Reinhilde Veugelers[4,5,6,7,*]

This PDF file includes:

- Materials, data processing & sample
    - Empirical scope of data base
    - Disambiguation of links to PubMed in patent data
    - Patent assignee information
    - Frontier science measurement
    - Sample selection considerations
- Additional results & robustness
    - Figures SM1 to SM4
    - Tables SM1 to SM21
- References

[1] Technical University of Munich, TUM School of Management, Dept. of Economics & Policy, Arcisstraße 21, D-80333 Munich, Germany. E-mail address: thomas.schaper@tum.de.
[2] KU Leuven, Faculty of Economics and Business, Dept. of Management, Strategy and Innovation, Naamsestraat 69, B-3000 Leuven, Belgium.
[3] KU Leuven, Faculty of Economics and Business Dept. of Management, Strategy and Innovation, Hendrik Conscienceplein 8, B-2000 Antwerp, Belgium. E-mail address: sam.arts@kuleuven.be.
[4] KU Leuven, Faculty of Economics and Business, Dept. of Management, Strategy and Innovation, Naamsestraat 69, B-3000 Leuven, Belgium. E-mail address: reinhilde.veugelers@kuleuven.be.
[5] Bruegel, Brussels, Belgium.
[6] PIIE, Washington, D.C., United States.
[7] CEPR, London, United Kingdom.
[*] *corresponding author.*

MATERIALS, DATA PROCESSING & SAMPLE

Empirical scope of data base

Our measurement of frontier author impact in technology draws, broadly, on the comprehensive population of all patents worldwide covered in the European Patent Office (EPO) PATSTAT Global database (autumn 2022 edition) and all scientific publications indexed in PubMed (until July 2018). For the reasons discussed below, we delimit our patent sample population to all biomedical U.S. patents assigned to private sector corporations granted by the United States Patent and Trademark Office (USPTO) between 1980–2009.

We define biomedical inventions based on three-digit USPC patent classes assigned to the 'Drugs&Medical' category in the classification of Hall et al. (2001).[1] Patents within this category are still a rather heterogeneous group, spanning from drug compounds (classes: 424, 435,514), to biotechnology and genetically modified organisms (classes: 800, 930), to surgery and medical instruments (classes: 128, 600, 601, 602, 604, 606, 607), to a miscellaneous group including optics (class: 351), dentistry (class: 433) and prothesis (class: 623). In our sample, we include all patents listing at least one such biomedical class. However, all results remain robust for the subset based on biomedical primary class only.

PubMed is a comprehensive online bibliographic index database for all, but not exclusively, biomedical literature, maintained by the National Centre of Biotechnology Information (NCBI) at the U.S. National Library of Medicine (NLM) (see: pubmed.ncbi.nlm.nih.gov). It currently covers citations to 32 million biomedical publications worldwide.

Disambiguation of links to PubMed in patent data

Inventor-author links. In order to identify inventors simultaneously (or previously) active in basic scientific discovery, we establish a link between all individuals listed as U.S. patent inventors (N = 2,665,709) between 1975 to 2011in the Li et al. (2014) database and all author identities in the *Author-ity* 2009 disambiguated database of author names (N = 9,243,805) in PubMed up until 2009 (Smalheiser &

[1] We use the 2006 version of the NBER classification.

Torvik, 2009; Torvik & Smalheiser, 2009). For this, we use the author-linked data provided by Torvik (2018) and match each author-patent instance to the series of all unique inventor identifiers in the same patent in Li et al. (2014). This results in an interim dataset, having author IDs (*au_id*) and author name strings on the left hand side and a vector of possible inventor IDs (*lower)* and their name strings on the right-hand side, which are connected to a joint patent number.

In a next step, we use standard within-patent probabilistic name matching techniques based on Jaro-Winkler similarity and Levinstein Distance across all author-patent-inventor instances in the data to determine the unique name pairs with the highest match probabilities and define a series of thresholds above which we accept a match to be unambiguous[2]: In Level 1 of this procedure, we accept only *unique* matches of author-inventor pairs with a Jaro-Winkler similarity score above 80% based on full name strings. In Level 2, out of the remaining, we accept all those pairs that score high (> 90%) on last name on both similarity measures, in case a match is unique above this threshold. In Level 3, finally, we accept those that yield unique matches on a threshold above 90% for Jaro-Winkler only, but slightly lower for Levinstein. This way, we are ultimately able to match 295,694 out of the 309,395 authors linked to a patent in the Torvik (2018) data (95,67%). Out of these 293,249 (99.08%) are uniquely disambiguated, 2,534 (0.86%) are matched to two Li et al. (2014) inventor IDs, 141 (0. 05%) are linked to three inventor IDs, and 40 (0. 01%) to four to a maximum of seven inventor IDs. We keep these very few cases of multiple matches in our sample, given their occurrence is unlikely to be correlated to any of the outcomes in our analyses. The remaining, very small, fraction of author-patent cases, for which we could not establish an automatic link, are virtually impossible to disambiguate even manually (usually very short or truncated name strings).

Next, we conduct a validation exercise of our matching approach. For this, we randomly sampled 200 candidates of inventor-author links (including null links, for which we do extensive web searches) from

[2] Jaro-Winkler similarity is a comparative measure yielding a score between 0 and 100 for the partial agreement between any two strings, taking into account the length of a string and also, in part, common human errors in spelling of names and number series. The Levinstein Distance is another commonly used similarity measure which calculates the number of string changes required from transforming one string into another. We use Jaro-Winkler as main algorithm, and Levinstein D to confirm accepted matches.

patents in the sample, and manually compare inventor name, location, assignee, co-inventors, narrow topic of patent to the matched author name, affiliation, location, co-authors, field of their (in time) closest publication(s) in PubMed. We further compare career patents' assignees, location, co-inventors to career publications' affiliations, locations, co-authors for matched pairs to assess their temporal co-evolution and plausibility. We find that, based on this procedure, 93.3% of inventor-author links are correctly identified (95.4% of positive links); 4.5% are very likely false positive author-patent links in the Torvik (2018) data. The remaining 2.2% are false negatives from our probabilistic name matching approach. In the validation set, no matched observations were false positives.

Scientific references (SNRPs). We use two data sources to identify references to scientific articles in U.S. patents: First, the dataset based on the *PatCi* algorithm (Agarwal, Lincoln, Cai, & Torvik, 2014), which disambiguates patent references to all articles in PubMed up until 2013. Second, we rely on the *PCS* dataset from Marx & Fuegi (2020), which provides references to all articles in Microsoft Academic Graph (MAG, Sinha et al. (2015)), incl. articles in PubMed.

We compare these two sources vis-à-vis with regards to their accuracy (precision and recall) for the patents in our sample: We extract two random samples of 100 SNPR links each (including null links), one based on *PatCi*, one based on *PCS*. We then manually check whether any reference listed (and disambiguated) on the matched USPTO patent's PDF document corresponds to the linked publication in PubMed online, in order to identify false positives. Second, we validate whether the other disambiguation finds the same match, to identify *conditional* false negatives. Finally, we repeat this procedure for the second extracted random sample. We find that both datasets have a precision of 100% (no false positive links) in the validation set. However, while all PubMed SNPRs of *PCS* are correctly found by *PatCi*,, *PCS* fails to detect about 22% of total references that are in *PatCi*. Given the importance of accurately identified references to PubMed for our analyses, we therefore use *PatCi* as our main SNPR disambiguation. However, we use *PCS* to replenish information on examiner added references (available for patents granted after January 2001), which is not available in *PatCi*.

Finally, we evaluate what the restriction on only SNPR to PubMed implies for generalizability and coverage of total scientific references made by patents in our sample. For this purpose, we compare the overlap between total references to scientific publications in PCS/ Microsoft Academic Graph (MAG) (spanning all scientific disciplines) and references specifically to PubMed indexed articles. This allows us to infer about how many links to non-biomedical articles supporting the inventions in our sample we may miss (i.e. not observe) in our data. We find this share of links to be very small and likely innocuous for internal validity of our estimates: In total, *PCS* identifies 1,879,926 unique SNPRs from patents in our sample to MAG. In contrast, with *PatCi* we observe only 1,617,074 unique SNPRs for the same patents to PubMed (86% of total MAG links). While 14% of missed links may still seem relatively considerable, we find this share to be significantly reduced when considering only references to (peer-reviewed) scientific journals indexed in the Science Citation Index (SCI): Our PubMed/ SCI SNPRs sample covers 94.9% of the total 1,492,540 references to scientific articles in SCI journals that are made by patents in our sample, implying that only 5% of links go to scientific articles outside of what we defined as the relevant population for our paper. This reinforces our belief of our results being representative for the universe of biomedical sciences.

Patent assignee information

We restrict our sample to patents assigned exclusively to private sector corporations based on the PATSTAT sector allocation (PSN), excluding collaborative patents between industry and universities or public research institutes. In order to identify individual firms, we rely on the U.S. patents' assignee disambiguation from PatentsView. In order to account for prior patenting of firms in the sample, we calculate firm patent stocks with a perpetual inventor method (Griliches, 1979; Hall, Jaffe, & Trajtenberg, Market value and patent citations, 2005), using a 15% annual depreciation rate. This allows us to discount firm R&D assets over time, despite not observing the yearly amount of expenditure. For instance, frontier authors may be particularly likely to appear on patents of small start-ups or spin-offs, which are characterize by high-growth rates and higher breakthrough likelihood (Schumpeter, 1942; Acs & Audretsch, 1990). On the other hand, large pharmaceutical firms more often have in-house research departments, closer ties to external research departments, and generate more valuable patents, on average (Henderson & Cockburn, 1996; Arora, Cohen, Lee, & Sebastian, 2023).

An important limitation of patent data is that, while we observe the sectoral allocation of patent applicants, we cannot consistently identify the affiliations of inventors at the time of patenting. A significant residual share of patent in our sample might be inventions resulting from informal or contractual collaborations of firms with individual, academic or public sector researchers.

Frontier science measurement

Scientific non-patent references. Our primary measurement of “frontier science” is based on recent (last three years) research articles published in top-general journal in biomedical sciences. The advantage of journal based indicators is that they allow for an “ex-ante” measurement of article quality provided by the peer-review selection, which is most likely unaffected by cumulative technological success of any patents associated with an article (either through an author- or SNPR-link). This helps minimizing concerns of simultaneity in measurement in our estimations.

However, we acknowledge that not all frontier research may be published in top-ranked journals and not all articles in such journals may entail frontier contributions. Therefore, we additionally use an alternative set of indicators to identify frontier science considering as such individual articles that enter the upper fifth percentile of the ten-year science citations distribution of their year-field cohort during the first three years after their publication (see Tables SM2 and SM10). In this case, to mitigate the simultaneity concerns discussed above, we exclude all patents from the estimation that we identify as “patent-paper-pairs” (Murray, 2002; Murray & Stern, 2007; Magerman, Van Looy, & Debackere, 2015). We detect “patent-paper-pairs” as patents that are linked to a scientific article through a non-patent reference, which is i) published in less than one year before or shortly after the patent application (enabling disclosure), and

ii) with whom there is a 100% overlap between inventors listed on the patent and authors on the publication, similar to the approach of Thompson et al. (2018).

For patent application dates, we consider the international priority date. In case of continuations or divisional filings, we consider the filing date of the application that ultimately led to the focal patent. For scientific articles, we consider the year of publication. PubMed collects the year of publication for each indexed article (i.e., the year in which the article has first been published in a peer-reviewed scientific journal). Exact dates of publication are unstructured and not systematically reported.

New scientific words in patent text. As an alternative measure to (frontier) SNPRs, we consider the reuse of new scientific words in patents' text bodies. Our basic approach is similar to the one of Iaria, Schwarz, & Waldinger (2018): we detect new words (i.e., new semantic terms) appearing for the first time, in the titles or abstracts of all 11,467,992 articles in PubMed in SCI indexed journals since 1878 until 2009. For this PubMed/SCI subset, title text is available for all articles, while abstract text is provided for 69.6% of articles. We lowercase all letters and exclude any signs and numbers from article text. We rely on words from both titles and abstracts in order to reduce the number of false negative, however our results are robust to only considering new words appearing in titles. To further minimize measurement error, we remove all 36,662 most frequent words in the English language (as identified by Project Gutenberg (2006)).[3] We additionally remove also regular conjugations and plural forms of words in the Project Gutenberg list (e.g., adding -ed, -ing, -ly, -s endings). We further only consider new words appearing after 1970, such that any frequent scientific terms, which may not be common in English per se, would not be mistakenly recorded as new, given they would have previously appeared in articles between 1878–1970. Given that it is difficult to establish the exact relative timing and priority of articles that were published in close temporal proximity to each other regarding their use of a new scientific term, and that such information would be little meaningful for our purposes, we consider words as new in all articles that mention them in the publication

[3] This list includes the most frequently used terms in English-language books in the Project Gutenberg as of April 2006. The full list is available at https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists#Project_Gutenberg.

year in which they first appear. Next, we introduce a quality dimension and categorize new scientific words by the highest journal rank in which they first appear, namely into “top-general”, “top-field”, and “non top-“ journal new words. In the case that words are mentioned in several articles and in different journals during the year of their introduction, we assign them to the category they appear in which has the relatively highest journal quality.

Subsequently, we trace the reuse of new scientific words in patents from the text body data of patent documents. For this, we consider reuses of new words that appear in the summary section of the invention in each of the 4,088,048 patents granted by the USPTO until 2009. The ‘invention summary’ section is typically a few hundred words long and reflects the essence of the technical description of the patented invention. For each patent, we only consider reuses of new words that had appeared in scientific articles latest by the time of patent grant.

In order to establish the frontier science dimension, we further categorize reuses of new words in patents by the recency since a word’s first appearance in a scientific article at the time of patent priority filing. Thereby, analogously to our SNPR measures, we consider as “frontier science words” new words from top-general journals reused in a patents within three years since their introduction, and “non-frontier science words” — respectively denoting non-recent top-general, top-field and non-top journal new word reuses.

This measure allows us to capture and qualify links to frontier (and non-frontier) science in patents through semantic analysis, independent of any references to scientific articles. Most patents in our sample include at least one new scientific word at some prior point in time, with new words in top-general journals reappearing most frequently (in 52% of patents). Frontier science words, however, are reused only in 5.3% of patents. Among the top-50 most frequent frontier science new words in patents in our sample are, by year of introduction in articles (new word year; number of reusing patents within first three years): mallophaga (1976; 105), IFNs (interferons 1980; 86), vaporariorum (1983; 139), LAV (1985; 86), integrins

(1987; 91), ribozymes (1988; 88), helicobacter (1989; 88), angiostatin (1994; 94), FasL (Fas Ligand 1994, 82), leptin (1995; 242), angiopoietin (1996; 91), caspase (1997; 404), epothilone (1997; 97), endostatin (1997; 216), siRNA (2001; 165).

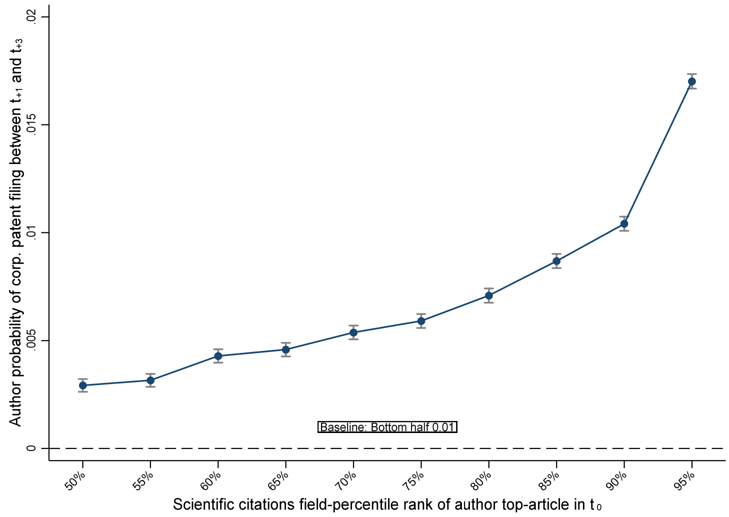
Sample selection considerations

Focus on biomedical research. Like in most of the related literature to our paper, the scope of our analysis is limited to biomedical research. This limitation arises, partly, due to the demarcation of our empirical data base. We can think of two ways in which this could possibly affect the generalizability of and external validity of our estimates: First, it could be that the more immediate and direct applicability of scientific knowledge makes the within-individual link particularly important in this area, leading us to potentially overvalue the role of frontier authors in technology when attempting to extrapolate to other areas. However, prior studies find that, irrespective of average science intensity, the relationship between direct science connectedness and technological impact — in particular between high-quality science and superior impact — is very stable and differs only slightly in magnitude across technological fields (Ahmadpoor & Jones, 2017; Poege, Harhoff, Gaessler, & Baruffaldi, 2019). This corroborates the view that frontier scientific insight may provide strong benefits for more useful applications, also in other areas. Second, we could also think about entry into patenting as a positive selection mechanism. While in biomedical research, engaging in technology development appears a comparatively small step from fundamental science, in other fields, where this is less common, participation in corporate patenting may imply a significantly higher selection hurdle, involving substantial risk and higher opportunity costs. Only scientists who can expect high gains and have a strong ability to translate their ideas into valuable applications, would then enter commercial technology development. In this case, finding a positive relationship in our biomedical patents would likely make us underestimate the importance of frontier authors, conditional on selection, for technological progress in other sectors. Third, independent of sorting of scientists into corporate patenting, the complementary skill-set required to translate scientific expertise into successful inventions might differs significantly more from the core skills of a scientist in some disciplines, compared to biomedical. In such cases, regardless of cognitive proximity and scientific aptitude, it is thinkable that we would overstate the impact of frontier scientists in technology in these disciplines. This is a limitation we cannot ultimately rule out, nor further evaluate, within the scope of our analyses and, thus, shall remain subject to further research.

Author selection into patents sample. In order to develop a better understanding regarding the selection mechanism governing participation in corporate technology of medical scientists, we estimate the probability of entering our inventor sample for the population of all authors in PubMed/ SCI publishing in a given year in the period 1980–2005 as a function of their highest ranked publication in that year in the citations distribution of a given field.

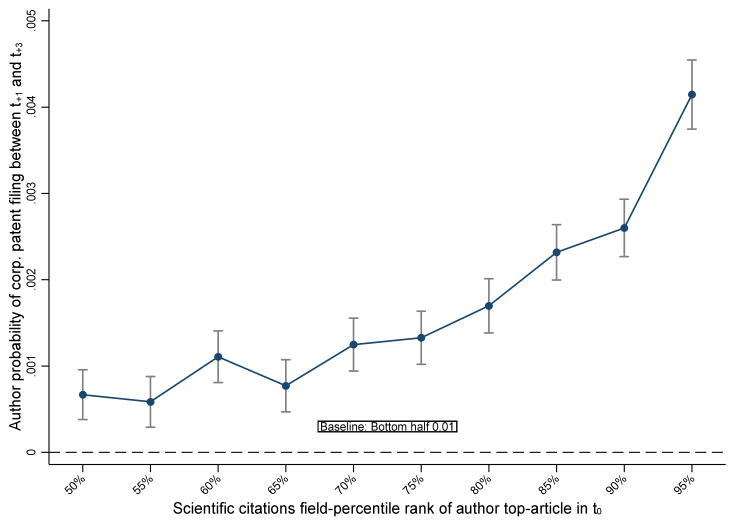
Figure SM1 plots the estimated relationship, after absorbing field by year idiosyncratic variation. We find the likelihood of appearing as inventor on a firm patent in our sample within three years after publication to be 1% for authors of scientific articles in the bottom half of the ten-year citations distribution. However, this probability steeply increases for authors of higher impact articles when moving towards the tail of the distribution. For authors of publications in the upper tenth percentile, the patenting probability is twice as high, for upper fifth percentile authors, it reaches +148% relative to the background rate. Even when comparing within same author variation over time (author fixed effects), the likelihood of appearing as inventor on a corporate patent increases by about +20% after authoring an upper tenth percentile most cited article and more than +40% after an upper fifth percentile article (Figure SM2). This relationship is robust for author-type classification based on the ranking of journals, as well as when comparing expected participation rates for within-author publishing variation (see Table SM1). This shows that authors of recent top-science tend to be more likely to become inventors, suggesting a strong and direct connect between creation of knowledge at the frontier and technological opportunity.

Figure SM1: Author-selection into industry patenting — across authors comparison



Notes: The figure plots parameter estimates from regressions of likelihood of entry into corporate patenting on frontier author-article citations percentile ranks for the yearly cross-section of all authors on articles in PubMed /SCI published between 1980–2005. The dependent variable measures the probability that an author appears as inventor on a U.S. patent assigned to a private sector corporation within three years following the publication year of the focal scientific article. The right-hand side / x-axis indexes percentiles of the ten-year citation distribution of a WOS subject category field year cohort. The estimation model includes a full set of publication year fixed effects. "Patent-paper pairs" are excluded. 95% confidence intervals are based on author-level clustered standard errors. For details on data sources and measures, see Section 3.

Figure SM2: Author-selection into industry patenting — within authors comparison



Notes: The figure plots parameter estimates from regressions of likelihood of entry into corporate patenting on frontier author-article citations percentile ranks for the yearly cross-section of all authors on articles in PubMed / SCI published between 1980–2005. The dependent variable measures the probability that an author appears as inventor on a U.S. patent assigned to a private sector corporation within three years following the publication year of the focal scientific article. The right-hand side / x-axis indexes percentiles of the ten-year citation distribution of a WOS subject category field year cohort. The estimation model includes a full set of author and publication year fixed effects. "Patent-paper pairs" are excluded. 95% confidence intervals are based on author-level clustered standard errors. For details on data sources and measures, see Section 3.

Table SM1: Author-selection into industry patenting

	OLS					
	Frontier = Top-general journal			*Frontier = Top-5% papers*		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	CorpPat3y	CorpPat3y	CorpPat3y	CorpPat3y	CorpPat3y	CorpPat3y
Frontier author	0.0083***	0.0011***	0.0010***	0.0158***	0.0034***	0.0031***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
WOS field x year	Yes	No	No	Yes	No	No
Year only	No	Yes	Yes	No	Yes	Yes
Author	No	Yes	Yes	No	Yes	Yes
Patent-paper pairs excluded	No	No	Yes	No	No	Yes
Observations	15,416,284	13,100,464	13,100,464	15,416,284	13,100,464	13,100,464
Adjusted R-squared	0.02	0.47	0.47	0.02	0.47	0.47

Notes: Each column reports parameter estimates from regressions of likelihood of entry into corporate patenting on frontier author status for the yearly cross-section of all authors on articles in PubMed / SCI published between 1980–2005. The dependent variable measures the probability that an author appears as inventor on a U.S. patent assigned to a private sector corporation within 3 years following the publication year of the focal scientific article. For details on data sources and measures, see Section 3. Standard errors in parentheses are heteroscedasticity robust in columns (1) and (4) and clustered at the author level in columns (2), (3), (5) and (6). Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM2: Detailed summary statistics

VARIABLES	Obs	Mean	SD	Median	25th	75th	90th	99th
Patent forward citations (10y)	237,345	17.28	34.76	6	2	18	43	166
Top-5% hit patent	237,345	0.07	0.25	0	0	0	0	1
$-value (in mio USD)	86,059	20.52	43.30	5	1	21	52	217
Number of inventors	237,345	3.00	2.12	2	2	4	6	11
Inventor-author	237,345	0.59	0.49	1	0	1	1	1
Number of inventor-authors	237,345	1.13	1.42	1	0	2	3	6
Frontier author (top-general journal, last 3y)	237,345	0.07	0.26	0	0	0	0	1
Frontier author (top-5% paper, last 3y)	237,345	0.12	0.32	0	0	0	1	1
Scientific reference (SNPR)	237,345	0.40	0.49	0	0	1	1	1
Number of SNPRs	237,345	4.81	15.91	0	0	3	12	70
Frontier SNPR (top-general journal, <= 3y)	237,345	0.09	0.28	0	0	0	0	1
Frontier SNPR (top-5% paper, <= 3y)	237,345	0.14	0.34	0	0	0	1	1
Citation lag top-general SNPR (years)	33,765	4.05	6.86	2	0	6	12	32
Citation lag top-5% SNPR (years)	47,630	3.02	5.64	2	0	5	9	25
Patent-paper pair	237,345	0.02	0.13	0	0	0	0	1
Inventor prior patents (max)	237,345	24.21	42.60	12	5	27	55	207
Inventor prior scientific publications (max)	237,345	11.10	23.70	2	0	11	31	120
Applicant firm patent stock	237,345	410.32	1,044.71	67	11	381	1,000	5,172

Notes: The table reports detailed patent-level summary statistics for the ample are all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). For details on data sources and measures, see Section 3.

Table SM3: Inventor-author-level summary statistics

VARIABLES	All	Frontier Authors	Non-frontier Authors	Non-Authors
		Prior patenting experience		
Number of prior USPTO patents	14.95	41.61	17.17	11.63
Inventor age (years since first patent)	6.64	6.87	8.14	5.60
Inventor network size (number of prior co-inventors)	12.91	18.65	15.42	10.81
		Prior scientific publishing experience		
Number of scientific articles	9.32	31.82	11.26	/
Number of first authored scientific articles	2.66	7.63	3.36	/
Number of total PubMed publications	12.61	38.7	15.46	/
Author age (years since first published article)	10.45	11.69	10.32	/
Years since last scientific article	2.04	0.23	2.22	/
Author network size (number of prior co-authors)	42.52	111.18	35.63	/
Author citations percentile rank in WOS-field	69.36	93.07	66.97	/
Inventor-patent instances	631,246	24,662	245,550	361,034
Unique individuals	159,994	6,457	48,913	104,624

Notes: The table reports inventor-patent-level group means. Sample are all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). “Frontier authors" are inventors that authored a paper in a "top-general" journal in the past three years before the patent application date. "Frontier SNPRs" are references to papers in "top-general" journals published in the same three years window (incl. articles published after first patent filing). Our scientific articles sample consists of all articles in PubMed indexed in Science Citation Index (SCI) journals. For details on data sources and measures, see Section 3.

Table SM4: Frontier author patents and technology impact — robustness

	PPML			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
Author	0.103***			0.017***		
	(0.009)			(0.001)		
Frontier author (top-general journal, last 3y)		0.124***	0.262***		0.015***	0.031***
		(0.019)	(0.022)		(0.002)	(0.003)
Author, top-general journal, > 3y			0.206***			0.024***
			(0.018)			(0.002)
Author, top field-journal, last 3y			0.160***			0.018***
			(0.015)			(0.002)
Author, top field-journal, > 3y			0.073***			0.009***
			(0.017)			(0.002)
Ln(inventor age)	Yes	Yes	Yes	Yes	Yes	Yes
Ln(number of prior patents)	Yes	Yes	Yes	Yes	Yes	Yes
Ln(author age)	No	Yes	Yes	No	Yes	Yes
Ln(number of prior articles)	No	Yes	Yes	No	Yes	Yes
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Number of authors	No	Yes	Yes	No	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	237,114	138,513	138,513	237,124	138,532	138,532
Pseudo/ adjusted R-squared	0.29	0.3	0.3	0.03	0.03	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM5: Frontier author patents and technology impact — firm fixed effects

	PPML			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
Author	0.055***			0.008***		
	(0.009)			(0.001)		
Frontier author (top-general journal, last 3y)		0.066***	0.069***		0.012***	0.013***
		(0.019)	(0.020)		(0.003)	(0.003)
Non-frontier author		0.054***			0.008***	
		(0.009)			(0.001)	
Author, top-general journal, > 3y			0.052***			0.009***
			(0.016)			(0.002)
Author, top field-journal, last 3y			0.082***			0.011***
			(0.013)			(0.002)
Author, top field-journal, > 3y			0.041***			0.005**
			(0.015)			(0.002)
Author, no top-journal			0.047***			0.007***
			(0.011)			(0.002)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Assignee firm	Yes	Yes	Yes	Yes	Yes	Yes
Observations	224,284	224,284	224,284	224,685	224,685	224,685
Pseudo/ adjusted R-squared	0.46	0.46	0.46	0.13	0.13	0.13

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of its primary patent class-year-citations distribution, in columns (4)-(6). Assignee fixed effects are based on PatentsView. For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table SM6: Frontier author patents and technology impact — "Biggest hits"

	OLS		
	(1)	(2)	(3)
VARIABLES	Biggest hit	Biggest hit	Biggest hit
Author	0.007***		
	(0.001)		
Frontier author (top-general journal, last 3y)		0.013***	0.014***
		(0.001)	(0.001)
Non-frontier author		0.006***	
		(0.001)	
Author, top-general journal, > 3y			0.009***
			(0.001)
Author, top field-journal, last 3y			0.008***
			(0.001)
Author, top field-journal, > 3y			0.006***
			(0.001)
Author, no top-journal			0.004***
			(0.001)
Patent class x year	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes
Observations	237,124	237,124	237,124
Pseudo/ adjusted R-squared	0.03	0.03	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures the probability that a patent enters the upper first percentile of the primary patent class-year-citations distribution. For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table SM7: Frontier author patents and patent value — private value indicators

	OLS								
	Kogan et al. values (in mio USD)			*Patent scope: Claim length*			*Patent renewal fees payments*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Ln($-value)	Ln($-value)	Ln($-value)	Ln(No. of wrds)	Ln(No. of wrds)	Ln(No. of wrds)	No. of renewals	No. of renewals	No. of renewals
Author	0.107***			-0.013***			0.000		
	(0.016)			(0.003)			(0.005)		
Frontier author (top-general journal, last 3y)		0.459***	0.493***		-0.094***	-0.092***		0.035***	0.036***
		(0.025)	(0.025)		(0.008)	(0.008)		(0.009)	(0.009)
Non-frontier author		0.068***			-0.005			-0.003	
		(0.016)			(0.004)			(0.005)	
Author, top-general journal, > 3y			0.216***			-0.052***			0.018**
			(0.024)			(0.007)			(0.008)
Author, top field-journal, last 3y			0.276***			0.101***			-0.027***
			(0.021)			(0.006)			(0.007)
Author, top field-journal, > 3y			0.102***			0.010			-0.017**
			(0.026)			(0.007)			(0.009)
Author, no top-journal			-0.129***			-0.045***			0.005
			(0.020)			(0.004)			(0.006)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	85,504	85,504	85,504	235,475	235,475	235,475	231,975	231,975	231,975
Adjusted R-squared	0.22	0.22	0.22	0.17	0.17	0.17	0.12	0.12	0.12

Notes: Each column reports parameter estimates from OLS regressions of patent value on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009. The dependent variable measures (ln) private values in millions of USD, deflated to 1982 dollars, from Kogan et al. (2017), in columns (1)-(3), the (ln) length (word count) of the first independent claim, in columns (4)-(6), and the number of times renewal fees were paid for the patent, in columns (7)-(8). Renewal fees are due four, eight and twelve years after patent grant. For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table SM8: Frontier author patents and technology impact — top 5% article authors

	PPML			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
Author	0.138***			0.020***		
	(0.008)			(0.001)		
Frontier author (top5% article, last 3y)		0.266***	0.269***		0.037***	0.038***
		(0.014)	(0.014)		(0.002)	(0.002)
Non-frontier author		0.116***			0.017***	
		(0.009)			(0.001)	
Author, top5% article, > 3y			0.206***			0.026***
			(0.015)			(0.002)
Author, no top5% article			0.100***			0.015***
			(0.009)			(0.001)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Patent-paper pairs excluded	Yes	Yes	Yes	Yes	Yes	Yes
Observations	232,931	232,931	232,931	232,941	232,941	232,941
Pseudo/ adjusted R-squared	0.29	0.29	0.29	0.03	0.03	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (4)-(6). "Top-5% articles" are scientific articles that enter the upper fifth percentile of a field-year cohort's citation distribution. For details on data sources and measures, see Section 3 and SM:Frontier science measurement. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM9: Frontier author patents and technology impact — first author splits

	PPML				OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
First author	0.145***			0.020***		
	(0.009)			(0.001)		
Author, never first author	0.121***			0.017***		
	(0.013)			(0.002)		
Frontier author, first author		0.350***	0.282***		0.047***	0.037***
		(0.028)	(0.027)		(0.004)	(0.004)
Frontier author, not first author		0.201***	0.100***		0.028***	0.014***
		(0.023)	(0.023)		(0.003)	(0.003)
Non-frontier author, first author		0.136***			0.019***	
		(0.009)			(0.001)	
Non-frontier author, not first author		0.114***			0.017***	
		(0.013)			(0.002)	
Author, top-general journal, > 3y, first author			0.048**			0.007**
			(0.021)			(0.003)
Author, top-general journal, > 3y, not first author			0.109***			0.011***
			(0.019)			(0.002)
Author, top field-journal, last 3y, first author			0.083***			0.010***
			(0.021)			(0.003)
Auhthor, top field-journal, last 3y, not first author			0.146***			0.019***
			(0.016)			(0.002)
Author, top field-journal, > 3y, first author			0.091***			0.012***
			(0.015)			(0.002)
Author, top field-journal, > 3y, not first author			0.087***			0.007***
			(0.015)			(0.002)
Author, no top-journal			0.064***			0.009***
			(0.010)			(0.001)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	237,114	237,114	237,114	237,124	237,124	237,124
Pseudo/ adjusted R-squared	0.28	0.28	0.28	0.03	0.03	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table SM10: Frontier author patents and technology impact — family-to-family citations

		PPML			OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	FamCit10	FamCit10	FamCit10	Hit patent	Hit patent	Hit patent
Author	0.168***			0.028***		
	(0.008)			(0.001)		
Frontier author (top-general journal, last 3y)		0.331***	0.344***		0.050***	0.052***
		(0.017)	(0.017)		(0.003)	(0.003)
Non-frontier author		0.155***			0.025***	
		(0.008)			(0.001)	
Author, top-general journal, > 3y			0.251***			0.040***
			(0.013)			(0.002)
Author, top field-journal, last 3y			0.256***			0.036***
			(0.013)			(0.002)
Author, top field-journal, > 3y			0.146***			0.022***
			(0.013)			(0.002)
Author, no top-journal			0.081***			0.016***
			(0.009)			(0.002)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	237,116	237,116	237,116	237,124	237,124	237,124
Pseudo/ adjusted R-squared	0.31	0.31	0.31	0.03	0.03	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent family citations (DOCDB-to-DOCDB level), in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of its primary patent class-year-family-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM11: Frontier author patents and technology impact — alternative impact time windows

	PPML			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit5	PatCit15	PatCitTot	Hit patent (5y)	Hit patent (15y)	Hit patent (total)
Frontier author (top-general journal, last 3y)	0.267***	0.241***	0.244***	0.036***	0.029***	0.028***
	(0.019)	(0.020)	(0.024)	(0.002)	(0.002)	(0.002)
Non-frontier author	0.127***	0.116***	0.095***	0.019***	0.017***	0.015***
	(0.009)	(0.009)	(0.010)	(0.001)	(0.001)	(0.001)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	237,025	237,116	237,116	237,124	237,124	237,124
Pseudo/ adjusted R-squared	0.21	0.32	0.34	0.02	0.04	0.04

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures five,fifteen, all-years window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of its primary patent class-year-citations distribution over a five, fifteen, all years window, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table SM12a: Frontier author patents and technology impact — excluding continuation filings

	PPML			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
Author	0.163***			0.023***		
	(0.010)			(0.001)		
Frontier author (top-general journal, last 3y)		0.382***	0.393***		0.051***	0.053***
		(0.025)	(0.025)		(0.004)	(0.004)
Non-frontier author		0.149***			0.021***	
		(0.010)			(0.001)	
Author, top-general journal, > 3y			0.274***			0.035***
			(0.018)			(0.003)
Author, top field-journal, last 3y			0.210***			0.028***
			(0.016)			(0.002)
Author, top field-journal, > 3y			0.161***			0.018***
			(0.019)			(0.003)
Author, no top-journal			0.085***			0.014***
			(0.012)			(0.002)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	164,551	164,551	164,551	164,565	164,565	164,565
Pseudo/ adjusted R-squared	0.3	0.3	0.3	0.04	0.04	0.04

Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009, excluding continuation filings (n = 164.889). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of its primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM12b: Frontier author patents and technological impact — continuation filings only

		PPML			OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
Author	0.128***			0.015***		
	(0.015)			(0.002)		
Frontier author (top-general journal, last 3y)		0.224***	0.233***		0.024***	0.025***
		(0.027)	(0.027)		(0.003)	(0.003)
Non-frontier author		0.119***			0.014***	
		(0.015)			(0.002)	
Author, top-general journal, > 3y			0.165***			0.018***
			(0.024)			(0.003)
Author, top field-journal, last 3y			0.183***			0.022***
			(0.023)			(0.003)
Author, top field-journal, > 3y			0.107***			0.014***
			(0.027)			(0.003)
Author, no top-journal			0.075***			0.007***
			(0.019)			(0.002)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	71,945	71,945	71,945	71,980	71,980	71,980
Pseudo/ adjusted R-squared	0.29	0.29	0.29	0.04	0.04	0.04

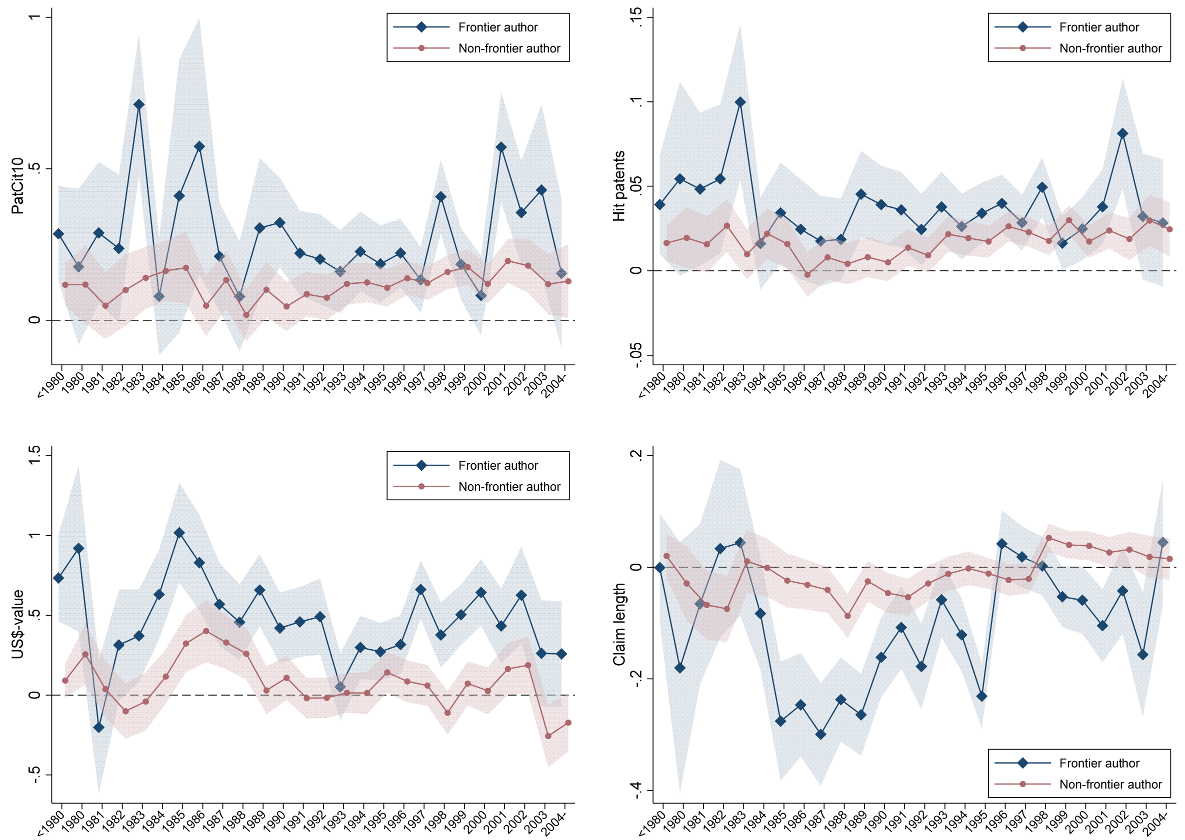
Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009, which are continuation filings (n = 72.456). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of its primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM13: Frontier author patents and technological impact — field splits

	PPML				OLS			
Technology fields	*Drugs*	*Medical instruments*	*Genetics*	*Other*	*Drugs*	*Medical instruments*	*Genetics*	*Other*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PatCit10	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent	Hit patent
Frontier author	0.291***	0.246***	0.320**	0.317***	0.034***	0.039***	0.068***	0.041*
	(0.023)	(0.040)	(0.145)	(0.093)	(0.003)	(0.008)	(0.018)	(0.022)
Non-frontier author	0.132***	0.120***	0.091	0.174***	0.016***	0.016***	0.028***	0.032***
	(0.013)	(0.012)	(0.130)	(0.029)	(0.002)	(0.002)	(0.009)	(0.005)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	121,250	67,247	5,106	13,259	121,250	67,247	5,106	13,259
Pseudo/ adjusted R-squared	0.06	0.12	0.08	0.28	0.01	0.01	0.01	0.01

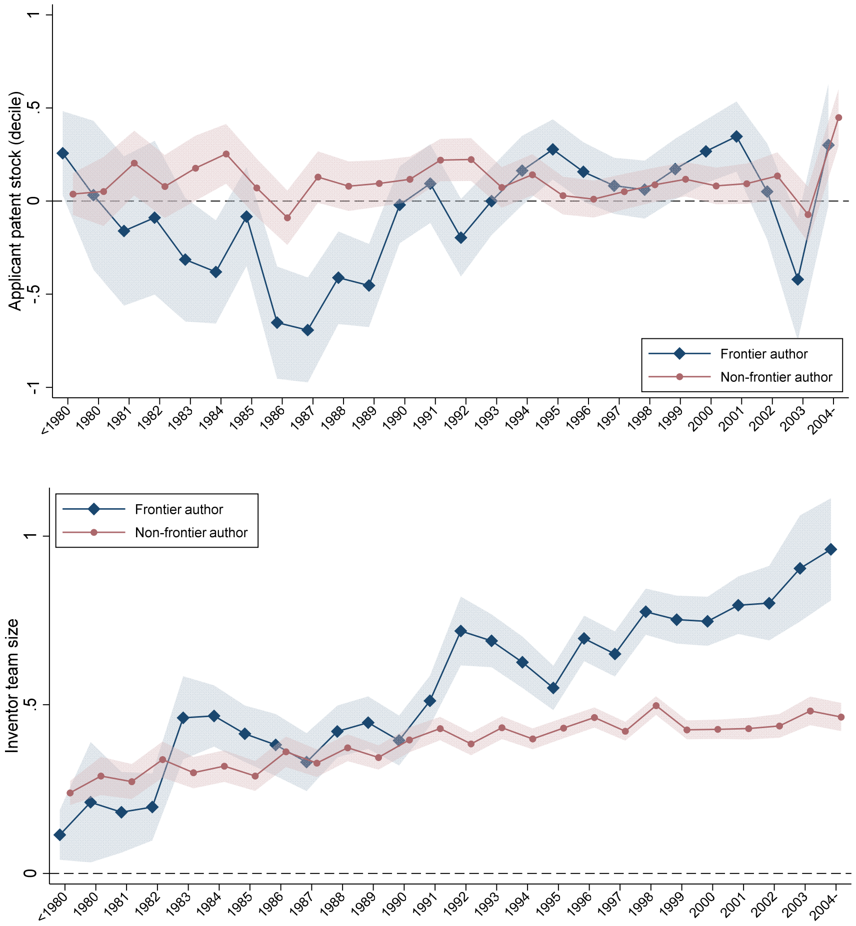
Notes: Each column reports parameter estimates from regressions of technology impact on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Figure SM3: Time trends in frontier author technology impact / private value premium



Notes: The figure plots parameter estimates for individual filing years from regressions of technology impact/ value on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in the upper left plot, the likelihood of entering the upper fifth percentile of the patent class-year-citations distribution, in the upper right plot, (ln) private values in millions of USD, deflated to 1982 dollars, from Kogan et al. (2017), in the lower left plot, and the (ln) length (word count) of the first independent claim, in the lower right plot. 95% confidence intervals are based on heteroscedasticity robust standard errors. All regressions control for class-by-year fixed effects, inventor team size and applicant patent stock characteristics. For details on data sources and measures, see Section 3.

Figure SM4: Time trends in assignee and inventor team size of frontier author patents



Notes: The figure plots parameter estimates for individual filing years from regressions of applicant firm size on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures the decile rank of assignee firms' patent stock at the time of patent filing, assuming a 15% stock depreciation rate, in the upper plot, and the (ln) number of co-inventors listed on a patent, in the bottom plot. 95% confidence intervals are based on heteroscedasticity robust standard errors. All regressions control for class-by-year fixed effects. The upper plot additionally controls for inventor team size, the bottom plot for assignee patent stock deciles & Δ. For details on data sources and measures, see Section 3.

Table SM14: Frontier SNPR patents and technology impact — top-SNPR recency splits

	PPML			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
Frontier SNPR (top-general journal, last 3y)	0.380***	0.350***	0.201***	0.051***	0.040***	0.024***
	(0.015)	(0.025)	(0.025)	(0.002)	(0.003)	(0.004)
SNPR, top-general journal, > 3y	0.281***	0.266***	0.138***	0.035***	0.027***	0.016***
	(0.019)	(0.025)	(0.025)	(0.003)	(0.003)	(0.004)
SNPR, top field-journal, last 3y	0.002	0.004	0.007	0.002	0.004	0.002
	(0.024)	(0.024)	(0.023)	(0.003)	(0.003)	(0.004)
SNPR, top field-journal, > 3y	-0.051**	-0.051**	-0.068***	-0.007**	-0.004	-0.007*
	(0.025)	(0.025)	(0.025)	(0.003)	(0.004)	(0.004)
SNPR, no top-journal	-0.059***	-0.059***	-0.005	-0.008***	-0.007***	0.001
	(0.020)	(0.021)	(0.019)	(0.003)	(0.003)	(0.003)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of SNPRs	No	Yes	Yes	No	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	No	Yes	Yes	No
Assignee firm	No	No	Yes	No	No	Yes
Observations	237,114	237,114	222,888	237,124	237,124	223,331
Pseudo / adjusted R-squared	0.29	0.29	0.47	0.03	0.03	0.14

Notes: Each column reports parameter estimates from regressions of technology impact on SNPR status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table SM15: Patents reusing frontier science words and technology impact

	PPML				OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PatCit10	PatCit10	PatCit10	Hit patent	Hit patent	Hit patent
Reuse of new scientific word	0.308***			0.033***		
	(0.024)			(0.002)		
Frontier science word (top-general journal, last 3y)		0.155***			0.018***	
		(0.036)			(0.004)	
Non-frontier science word		0.070**			0.005	
		(0.031)			(0.004)	
Frontier science word, first			0.022			0.003
			(0.071)			(0.008)
Frontier science word, not first			0.128***			0.016***
			(0.037)			(0.005)
Non-frontier science word, first			-0.015			-0.000
			(0.033)			(0.004)
Non-frontier science word, not first			0.069**			0.005
			(0.031)			(0.004)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of new scientific words	No	Yes	Yes	No	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	237,114	237,114	237,114	237,124	237,124	237,124
Pseudo / adjusted R-squared	0.28	0.29	0.29	0.03	0.03	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on reuse of new scientific words s for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1)-(3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (4)-(6). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM16: Frontier SNPR patents and patent value — private value indicators

	OLS								
	Kogan et al. values (in mio USD)			*Patent scope: Claim length*			*Patent renewal fees payments*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Ln($-value)	Ln($-value)	Ln($-value)	Ln(No. of wrds)	Ln(No. of wrds)	Ln(No. of wrds)	No. of renewals	No. of renewals	No. of renewals
SNPR	0.178***			-0.076***			0.030***		
	(0.015)			(0.004)			(0.005)		
Frontier SNPR (top-general journal, last 3y)		0.207***			-0.107***			0.030**	
		(0.033)			(0.010)			(0.012)	
Non-frontier SNPR		0.058**			-0.052***			0.026***	
		(0.025)			(0.007)			(0.008)	
Frontier SNPR, first			0.268***			-0.186***			0.093***
			(0.044)			(0.014)			(0.017)
Frontier SNPR, not first			0.185***			-0.101***			0.033**
			(0.036)			(0.011)			(0.013)
Non-frontier SNPR, first			0.058**			-0.087***			0.067***
			(0.029)			(0.008)			(0.010)
Non-frontier SNPR, not first			0.058**			-0.035***			0.008
			(0.026)			(0.007)			(0.009)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of SNPRs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	85,504	85,504	85,504	235,475	235,475	235,475	231,975	231,975	231,975
Adjusted R-squared	0.22	0.22	0.22	0.16	0.16	0.16	0.12	0.12	0.12

Notes: Each column reports parameter estimates from OLS regressions of patent value on SNPR status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009. The dependent variable measures (ln) private values in millions of USD, deflated to 1982 dollars, from Kogan et al. (2017), in columns (1)-(3), the (ln) length (word count) of the first independent claim, in columns (4)-(6), and the number of times renewal fees were paid for the patent, in columns (7)-(8). Renewal fees are due four, eight and twelve years after patent grant. For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table SM17: Frontier author patents and reuse of frontier science words

	OLS	PPML		OLS	
		with new words subsample		with author subsample	
	(1)	(2)	(3)	(4)	(5)
VARIABLES	New word	# new words	Frontier science word	First Frontier word	No self-reuse Frontier word
Frontier author	0.010***	0.189***	0.068***	0.048***	0.054***
	(0.001)	(0.012)	(0.003)	(0.009)	(0.003)
Non-frontier author	0.007***	0.109***	-0.001		
	(0.001)	(0.007)	(0.001)		
Constant	0.968***	3.655***	0.049***	0.153***	0.053***
	(0.001)	(0.006)	(0.001)	(0.017)	(0.003)
With frontier science word subsample				Yes	No
Author-level controls				Yes	Yes
DV subsample average	0.972	30.789	0.054	0.104	0.063
Patent class x year	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes
Number of new scientific words	No	No	Yes	Yes	Yes
Assignee firm	Yes	Yes	Yes	Yes	Yes
Observations	237,124	230,512	230,512	8,721	135,262
Adjusted / pseudo R-squared	0.02	0.40	0.12	0.08	0.11

Notes: Each column reports parameter estimates from regressions of reuse of new scientific words on inventor-author status for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures the probability that a patent reuses a new (frontier) scientific word appearing in articles indexed in PubMed / SCI, in columns (1) and (3)-(5), and the number of new words in column (2). Additional author-level controls in columns (4)-(5) are the logs of prior publications, patents, co-author network size, and number of authors fixed effects. For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM18: Frontier author patents, frontier SNPRs and patent value — private value indicators

	OLS					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Ln($-value)	Ln(Claim length)	No. of renewals	Ln($-value)	Ln(Claim length)	No. of renewals
Frontier author, SNPR	0.388***	-0.132***	0.062***			
	(0.036)	(0.011)	(0.013)			
Frontier author, no SNPR	0.479***	-0.037**	-0.002	0.479***	-0.037**	-0.003
	(0.050)	(0.015)	(0.018)	(0.050)	(0.015)	(0.018)
Non-frontier author, SNPR	0.100***	-0.040***	0.016*			
	(0.029)	(0.007)	(0.009)			
Non-frontier author, no SNPR	0.008	-0.008*	-0.003	0.008	-0.008*	-0.003
	(0.020)	(0.004)	(0.006)	(0.020)	(0.004)	(0.006)
Non-author, SNPR	0.017	-0.083***	0.035***			
	(0.030)	(0.008)	(0.010)			
Frontier author, frontier SNPR				0.498***	-0.165***	0.099***
				(0.043)	(0.014)	(0.017)
Frontier author, non-frontier SNPR				0.327***	-0.125***	0.019
				(0.042)	(0.014)	(0.016)
Non-frontier author, frontier SNPR				0.186***	-0.092***	-0.005
				(0.039)	(0.012)	(0.014)
Non-frontier author, non-frontier SNPR				0.100***	-0.037***	0.019**
				(0.029)	(0.007)	(0.010)
No author, frontier SNPR				0.068	-0.099***	0.005
				(0.059)	(0.017)	(0.022)
No author, non-frontier SNPR				0.023	-0.087***	0.037***
				(0.031)	(0.008)	(0.010)
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Number of SNPRs	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	85,504	235,475	231,975	85,504	235,475	231,975
Adjusted R-squared	0.22	0.17	0.12	0.22	0.17	0.12

Notes: Each column reports parameter estimates from regressions of private patent value on decompositions of (frontier) inventor-author status by (frontier) SNPR-type for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures (ln) private values in millions of USD, deflated to 1982 dollars, from Kogan et al. (2017), in columns (1) & (4), the (ln) length (word count) of the first independent claim, in columns (2) & (5), and the number of times renewal fees were paid, in columns (3) & (6). Renewal fees are due four, eight and twelve years after patent grant. For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table SM19: Frontier author patents, reuse of frontier science words and technology impact

	PPML	OLS	PPML	OLS
	(1)	(2)	(3)	(4)
VARIABLES	PatCit10	Hit patent	PatCit10	Hit patent
Frontier author, new scientific word	0.320***	0.037***		
	(0.040)	(0.005)		
Frontier author, no new scientific word	0.132	0.017*	0.132	0.017*
	(0.141)	(0.010)	(0.141)	(0.010)
Non-frontier author, new scientific word	0.203***	0.022***		
	(0.037)	(0.004)		
Non-frontier author, no new scientific word	0.166***	0.016***	0.165***	0.016***
	(0.048)	(0.005)	(0.048)	(0.005)
Non-author, new scientific word	0.098***	0.006		
	(0.036)	(0.004)		
Frontier author, frontier science word			0.272***	0.032***
			(0.053)	(0.007)
Frontier author, non-frontier science word			0.333***	0.039***
			(0.041)	(0.005)
Non-frontier author, frontier science word			0.302***	0.034***
			(0.045)	(0.006)
Non-frontier author, non-frontier science word			0.203***	0.022***
			(0.037)	(0.004)
No author, frontier science word			0.211***	0.025***
			(0.046)	(0.006)
No author, non-frontier science word			0.098***	0.006
			(0.036)	(0.004)
Patent class x year	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes
Number of new scientific words	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes
Observations	237,114	237,124	237,114	237,124
Pseudo / adjusted R-squared	0.29	0.03	0.29	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on decompositions of (frontier) inventor-author status by reuse of (frontier) scientific words for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in columns (1) and (3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-citations distribution, in columns (2) and (4). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table SM20: Frontier author patents and technology impact – self-citations & timing of frontier SNPRs

	PPML	OLS	PPML	OLS
	(1)	(2)	(3)	(4)
VARIABLES	PatCit10	Hit patent	PatCit10	Hit patent
Frontier author, frontier SNPR, self-citation	0.488***	0.049***		
	(0.050)	(0.007)		
Frontier author, frontier SNPR, no self-citation	0.334***	0.043***		
	(0.039)	(0.005)		
Non-frontier author, frontier SNPR	0.356***	0.045***		
	(0.027)	(0.004)		
No author, frontier SNPR	0.366***	0.042***		
	(0.041)	(0.006)		
Frontier author, frontier SNPR, first			0.267***	0.032***
			(0.051)	(0.006)
Frontier author, frontier SNPR, not first			0.466***	0.054***
			(0.039)	(0.005)
Non-frontier author, frontier SNPR, first			0.285***	0.037***
			(0.038)	(0.006)
Non-frontier author, frontier SNPR, not first			0.386***	0.049***
			(0.028)	(0.004)
No author, frontier SNPR, first			0.252***	0.036***
			(0.068)	(0.010)
No author, frontier SNPR, not first			0.419***	0.045***
			(0.047)	(0.007)
Author non-frontier & no SNPR interactions	Yes	Yes	Yes	Yes
No author non-frontier SNPR interactions	Yes	Yes	Yes	Yes
Patent class x year	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes
Number of SNPRs	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes
Observations	234,863	234,873	237,114	237,124
Pseudo / adjusted R-squared	0.29	0.03	0.29	0.03

Notes: Each column reports parameter estimates from regressions of technology impact on decompositions of (frontier) inventor-author status by (frontier) SNPR-type for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures ten-year window forward patent citations, in column (1) and (3), and the probability that a patent enters the upper fifth percentile of the primary patent class-year-impact distribution, in column (2) and (4). For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table SM21: Frontier author patents and (private) patent value – self-citations & timing of frontier SNPRs

	OLS					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Ln($-value)	Ln(Claim length)	No. of renewals	Ln($-value)	Ln(Claim length)	No. of renewals
Frontier author, frontier SNPR, self-citation	0.571***	-0.125***	0.148***			
	(0.057)	(0.020)	(0.024)			
Frontier author, frontier SNPR, no self-citation	0.459***	-0.187***	0.069***			
	(0.048)	(0.016)	(0.019)			
Non-frontier author, frontier SNPR	0.184***	-0.093***	-0.007			
	(0.039)	(0.012)	(0.014)			
No author, frontier SNPR	0.063	-0.099***	0.003			
	(0.059)	(0.017)	(0.022)			
Frontier author, frontier SNPR, first				0.494***	-0.208***	-0.208***
				(0.059)	(0.019)	(0.019)
Frontier author, frontier SNPR, not first				0.505***	-0.145***	-0.145***
				(0.047)	(0.016)	(0.016)
Non-frontier author, frontier SNPR, first				0.284***	-0.152***	-0.152***
				(0.054)	(0.017)	(0.017)
Non-frontier author, frontier SNPR, not first				0.152***	-0.073***	-0.073***
				(0.042)	(0.013)	(0.013)
No author, frontier SNPR, first				0.046	-0.180***	-0.180***
				(0.100)	(0.027)	(0.027)
No author, frontier SNPR, not first				0.080	-0.061***	-0.061***
				(0.067)	(0.021)	(0.021)
Author non-frontier & no SNPR interactions	Yes	Yes	Yes	Yes	Yes	Yes
No author non-frontier SNPR interactions	Yes	Yes	Yes	Yes	Yes	Yes
Patent class x year	Yes	Yes	Yes	Yes	Yes	Yes
Number of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Number of SNPRs	Yes	Yes	Yes	Yes	Yes	Yes
Firm patent stock decile x Δ	Yes	Yes	Yes	Yes	Yes	Yes
Observations	84,689	233,229	229,740	85,504	235,475	235,475
Adjusted R-squared	0.22	0.17	0.12	0.22	0.17	0.12

Notes: Each column reports parameter estimates from regressions of private patent value on decompositions of (frontier) inventor-author status by (frontier) SNPR-type for the sample of all biomedical U.S. patents exclusively assigned to firms and granted between 1980-2009 (N = 237,345). The dependent variable measures (ln) private values in millions of USD, deflated to 1982 dollars, from Kogan et al. 2017, in columns (1) & (4), the (ln) length (word count) of the first independent claim, in columns (2) & (5), and the number of times renewal fees were paid, in columns (3) & (6). Renewal fees are due four, eight and twelve years after patent grant. For details on data sources and measures, see Section 3. Heteroscedasticity robust standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01

REFERENCES

Acs, Z. J., & Audretsch, D. B. (1990). *Innovation and small firms.* Cambridge: MIT Press.

Agarwal, S., Lincoln, M., Cai, H., & Torvik, V. I. (2014). Patci—a tool for identifying scientific articles cited by patents.

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science, 357*, 583–587.

Arora, A., Cohen, W., Lee, H., & Sebastian, D. (2023). Invention value, inventive capability and the large firm advantage. *Research Policy, 52*(1), 1046-1050.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, 92-116.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. *The NBER patent citation data file: Lessons, insights and methodological tools*. National Bureau of Economic Research Cambridge, Mass., USA.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of economics*, 16-38.

Henderson, R., & Cockburn, I. (1996). Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *The RAND Journal of Economics*, 32-59.

Iaria, A., Schwarz, C., & Waldinger, F. (2018). Frontier knowledge and scientific production: evidence from the collapse of international science. *The Quarterly Journal of Economics, 133*(12), 927–991.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics, 132*(2), 665-712.

Li, G.-C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., . . . Fleming, L. (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). *Research Policy, 43*, 941–955.

Magerman, T., Van Looy, B., & Debackere, K. (2015). Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology. *Research Policy, 44*(9), 1702-1713.

Marx, M., & Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal, 41*(9), 1572-1594.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. *Research policy, 31*, 1389–1403.

Murray, F., & Stern, S. (2007). Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior & Organization, 63*, 648–687.

Poege, F., Harhoff, D., Gaessler, F., & Baruffaldi, S. (2019). Science quality and the value of inventions. *Science advances, 5*, eaay7323.

Schumpeter, J. A. (1942). *Socialism, capitalism and democracy.* Harper and Brothers.

Sinha, A. S., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An Overview of Microsoft Academic Service (MAS) and Applications. *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)* (pp. 243-246). New York, NY, USA: ACM.

Smalheiser, N. R., & Torvik, V. I. (2009). Author name disambiguation. *Annual review of information science and technology, 43*, 1–43.

Thompson, N. C., Ziedonis, A. A., & Mowery, D. C. (2018). University licensing and the flow of scientific knowledge. *Research Policy, 47*(6), 1060-1069.

Torvik, V. I. (2018). Author-Linked data for Author-ity 2009. *Author-Linked data for Author-ity 2009*. University of Illinois at Urbana-Champaign. doi:10.13012/B2IDB-4370459_V1

Torvik, V. I., & Smalheiser, N. R. (2009). Author name disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data (TKDD), 3*, 1–29.

KU LEUVEN
Faculty of Economics and Business
Management, Strategy and Innovation (MSI)
Naamsestraat 69 bus 3535
3000 LEUVEN, Belgium
tel. + 32 16 32 67 00
msi@econ.kuleuven.be
https://feb.kuleuven.be/research/MSI/