Scientific Talents and Firm Growth: Evidence from Scientific Breakthroughs

Jialin Qian*

July 2024

This paper studies the impact of corporate scientists on firm growth in the context of scientific breakthroughs. Leveraging a comprehensive publication database and a text-embedding large language processing tool, I develop a measure for corporate scientific human capital. Analysis of three major scientific breakthroughs of the 21st century reveals that firms affected by these breakthroughs and possessing a higher stock of relevant scientific human capital demonstrate superior performance following these breakthroughs. Corporate scientists create value mainly through the knowledge spillover channel. Specifically, corporate scientists engage more in patenting after scientific breakthroughs in firms affected by the breakthroughs and endowed with substantial scientific human capital. These firms also generate a greater number of impactful patents and are more likely to be early adopters of citing scientific papers in their patents compared to their peers. Additionally, these firms are more successful in attracting star scientists in the aftermath of breakthroughs. This study highlights the crucial role of corporate scientists in connecting basic science with industrial innovation in a modern economy that increasingly relies on intangible assets and human capital.

JEL Classification: C81;G3;J24;O3

Keywords: Corporate Innovation, Human Capital, Firm Growth, Corporate Policy, Corporate Scientists, Scientific Breakthroughs, Knowledge spillover, Textual analysis

*J. Mack Robinson College of Business, Georgia State University, Email: jqian5@gsu.edu

1. Introduction

Technological innovation is a crucial driver of long-term growth and value creation (Aghion and Howitt, 1990; Kogan et al., 2017). The 21st century has witnessed several milestones in the technological revolution that significantly impact human life and work. The COVID-19 mRNA vaccine marks the first successful large-scale use of mRNA in vaccine production, playing a pivotal role in addressing the crisis. The rapid advancement of artificial intelligence also presents numerous new opportunities for our future. The provision of new scientific knowledge is a fundamental source of novel ideas for industrial innovation (Sorenson and Fleming, 2004; Ahmadpoor and Jones, 2017; Krieger, Schnitzer, and Watzinger, 2022). In recent decades, university science has increased considerably, while firms have reduced their investment in upstream research, focusing more on the development and commercialization of scientific discoveries (Arora, Belenzon, and Sheer, 2021; Arora et al., 2021). Translating academic science into commercialized innovative products requires the ability to absorb knowledge embedded in academic research (Cockburn and Henderson, 1998; Rosenberg, 2010). In this new innovation ecosystem, corporate scientists play a crucial bridging role in fostering industrial inventiveness by absorbing knowledge from academic science while pursuing corporate innovation goals. Therefore, the importance of corporate scientists in driving firm value in the modern innovation system remains a question of great interest to both scholars and policymakers.

Despite the importance of this topic, there are limited studies that comprehensively examine how firms' investments in scientific human capital (shortened as *SHC*) drive firm growth. One reason for this gap is the challenge of measuring corporate scientific human capital (SHC). To address this, I leverage a publication database and a text-embedding large language processing tool to construct a measure of *SHC* for public firms in the U.S. I investigate whether scientific human capital contributes to firm value and innovation, and if so, how it plays a role.

I utilize exogenous shocks from scientific breakthroughs in public science within a stacked difference-in-differences framework. This approach allows me to leverage positive shocks to the innovation opportunities encountered by corporate scientists, highlighting their role in driving firm value. The exogenous shocks include three major 21st-century scientific breakthroughs: Human Genome Project; Deep Learning; and Gene Editing. These events were selected based on an editorial article in *Nature*[1], which summarizes groundbreaking innovations of the 21st century. The publication years of key academic papers behind these breakthroughs are used as the breakthrough years. These three scientific breakthroughs, originating from academia, are largely unforeseen by companies, thus presenting increased positive innovation and investment opportunities for affected firms while having no immediate impact on unaffected companies.

I classify firms into treated and control groups based on their significant exposure to a given scientific breakthrough. To measure the level of impact, I examine the most valuable patents of each company in the three years preceding the breakthrough. The patent value is determined using the method outlined by Kogan et al. (2017). The relevance of these patents to the scientific breakthrough is assessed by calculating the cosine similarity between the patent abstracts and the representative paper abstract of the scientific breakthrough, utilizing a text-embedding large language processing tool. Each patent is assigned a relevance score, and a company's relevance to each event is determined by the average similarity score of its most valuable patents prior to the event. Among all firms with patents filed within three years before the breakthrough year, the top 20% are classified as treated firms for a given event, while the remaining firms serve as control firms.

Firms whose core technology areas are related to these scientific breakthroughs are expected to experience enhanced performance post-breakthrough compared to those focusing on unrelated technology areas. Using a stacked difference-in-difference approach, I find that treated firms exhibit better operating profitability, and increased sales in the five years

[1] See https://www.nature.com/articles/d41586-023-04021-2

following the occurrence of a scientific breakthrough compared to control firms. Given that the three breakthrough events occur at different times, the stacked Difference-in-Differences (DID) approach mitigates concerns about biased estimates due to treatment effect heterogeneity, as raised in the traditional Difference-in-Differences setting (Baker, Larcker, and Wang, 2022). Moreover, the effect is economically significant. For example, treated firms' operating profitability increases by 2.7% more compared to control firms. Pre-trend tests reveal no significant pre-existing trends, confirming that the three scientific breakthroughs have a positive impact on relevant companies.

After identifying the affected firms, I examine the core question: what role does scientific human capital play in driving innovation and long-term growth among these firms? I expect that a treated firm's performance is better when it possesses a higher level of scientific human capital prior to the scientific breakthroughs. The prediction is based on the idea that firms with more scientific human capital have a higher absorptive ability for the cutting-edge knowledge embedded in the breakthroughs (Cockburn and Henderson, 1998; Rosenberg, 2010), thereby transferring the knowledge to their products in a more effective and efficient way. To explore the question, I construct a measure of corporate scientific human capital. From the *OpenAlex* database[2], I collect papers published by corporate scientists and their affiliations. Corporate scientists are defined as those who have published academic articles affiliated with a company. The primary scientific human capital used in the analysis is based on the historical publication record of employees affiliated with a firm. It is constructed as follows: first, the annual publication stock is obtained by counting the total number of scientific papers published by all employees affiliated with firm i in year t up to year t; second, the scientific human capital measure SHC is calculated by summing the annual publication stock within the window $[-3, -1]$ centered on the breakthrough year. To test robustness, two additional measures are constructed: corporate scientists affiliated with the firm six to

[2]OpenAlex offers an increasingly widely used industry-standard scientific knowledge base, covering 258 million papers up to 2024. https://docs.openalex.org/.

eight years before the shock, and the closeness of scientists' expertise to the breakthrough, using textual information from publication abstracts and the abstract of the paper signifying the breakthrough.

In a stacked triple difference-in-differences (DID) setting, the findings reveal that firms with higher levels of scientific human capital experience greater profitability and market value in the five years following scientific breakthroughs compared to other firms. To address the potential concern that investment in scientific human capital may be correlated with firm size, I control for firm size in all specifications. The results remain robust with control variables and multiple layers of fixed effects, underscoring the critical role of corporate scientific human capital in driving firm growth post-scientific breakthroughs.

Next, I conduct several tests to examine the channels through which scientific human capital impacts firm value post-breakthrough. The first is the knowledge spillover effect, where companies with high scientific human capital quickly absorb new knowledge (Cockburn and Henderson, 1998; Rosenberg, 2010). Scientists in these companies are expected to engage in more patenting activities and collaborate with inventors, fostering new ideas and producing impactful patents. The findings support this channel. First, I show that firms with a high stock of scientific human capital publish more impactful papers after the scientific breakthrough than peer firms, indicating a faster accumulation of scientific human capital and more improvement in absorptive ability to the new scientific knowledge embedded in scientific breakthroughs. Secondly, I show that treated firms with high scientific human capital have a higher proportion of scientist-inventors in their patents, produce higher quality patents, and are more likely to be among the first to cite the knowledge related to the scientific breakthroughs. These results provide evidence of the role of corporate scientists in increasing firm value by facilitating the transfer of scientific knowledge into the innovation process.

The second channel is the attraction effect for star scientists. According to anecdotal evidence and the findings of Ahmed (2022), support for publication is a key factor for scientists joining a company. A firm's stock of scientific human capital can be a proxy for its culture of supporting scientific publications and basic research. Therefore, I conjecture that companies with a higher stock of scientific human capital are more attractive to star scientists. My results confirm this hypothesis, showing that treated firms with higher levels of scientific human capital prior to scientific breakthroughs attract more star scientists post-breakthrough than their peer firms. This attraction effect complements the spillover effect, further strengthening firms' ability to absorb and transfer new scientific knowledge into better innovation outcomes.

The study contributes to labor and finance literature that highlights the role played by skilled labor in fostering firm growth and value creation. Existing studies mainly focus on general skilled labor, such as employees with STEM majors or advanced skills. For example, Shen (2021) utilizes the friction of skilled labor mobility during the Green Card application process to study the effect of skilled labor mobility on firm valuation. Babina et al. (2024) exploits the job posting and resume data to identify firms' demands on different skills as a proxy for AI investment and study its impact on firm growth. This study focuses on a special group of high-skill labor, corporate scientists, to examine the impact of top-tier scientific human capital in fostering corporate growth.

This paper also contributes to the understanding of corporate innovation in an economcy that increasingly relies on human capital and intangible capital. Previous research has examined various factors influencing corporate innovation, including corporate governance (Aghion, Van Reenen, and Zingales, 2013; Bena and Li, 2014; Brav et al., 2018), regulatory changes (Acharya, Baghai, and Subramanian, 2014,), financing policies (Acharya and Xu, 2017; Azoulay et al., 2019), and corporate culture (Li et al., 2021; Tian and Wang, 2014; Derrien, Kecskés, and Nguyen, 2023). This study adds to the strain of literature examining the effect of corporate culture on innovation by providing evidence that a culture that supports corporate research and scientific publication promotes corporate innovation, especially during scientific breakthroughs.

Lastly, I make a data contribution by using bibliometric and textual analysis to propose a new scientific human capital measure by identifying corporate scientists and connecting their publications with scientific breakthroughs. Previous studies have used bibliometric data to measure corporate investment in research (Arora, Belenzon, and Sheer, 2021; Arora et al., 2021), link university publications with patenting (Babina et al., 2023; Myers and Lanahan, 2022), A recent study (Arora et al., 2023) constructs a measure of firms' exposure to public science by connecting university publications, PhD dissertations, and firm patenting. There are some similarities between this paper and my study, as we both explore the effect of public science on industrial innovation output. However, there are some key differences between the two studies: First, my research focuses on highlighting the role of corporate scientists in driving firm innovation in response to unexpected scientific breakthroughs, whereas Arora et al. (2023) primarily examine how firms react to shocks from public science. Second, I concentrate exclusively on prominent scientific revolutions in the 21st century, while their study covers more minor scientific discoveries across various sub-fields. By identifying corporate scientists and tracking their publication and patenting activities, I am able to investigate the underexplored roles of corporate scientists in advancing firm growth and innovation.

2. Empirical strategy and data

2.1. Scientific breakthrough

Studying the role of scientific human capital in propelling firm growth at the time of scientific breakthroughs requires shocks to firms' exposure to these breakthroughs. I select the most prominent scientific breakthroughs in the 21st century as documented in an editorial article in *Nature*[3]. The article refers to the most extraordinary events of scientific disruption in the past two decades, including the Human Genome Project, the discovery of the Higgs

[3] See https://www.nature.com/articles/d41586-023-04021-2.

boson, gene editing and CRISPR technology, the first detection of gravitational waves, and AI and machine learning. Among these scientific breakthrough events, the discovery of the Higgs boson and the detection of gravitational waves belong to the field of theoretical physics and do not have a direct impact on industrial innovation on a large scale soon. Therefore, I keep the rest of the three events as the shocks of scientific breakthroughs throughout the study. A shared characteristic of the three remaining events is that they are all initiated by researchers in public research institutions instead of the private sector. They can serve as a quasi-exogenous shock to firms as firms cannot predict the timing of the breakthrough ex ante.

I define the breakthrough year of the three events as the year the paper that signifies the breakthrough is published. According to the editorial article in *Nature*, the accomplishment of the Human Genome Project is signified by the paper titled "Initial sequencing and analysis of the human genome" published in *Science* in 2001, a landmark paper that presented the draft sequence of the human genome, organized by the International Human Genome Sequencing Consortium. The second event, Deep Learning and Neural Networks, is marked by the finding in the paper "A fast learning algorithm for deep belief nets", published in *Neural Computation* in 2006 by one of the "Godfathers of AI," Geoffrey E. Hinton and co-authors. The paper was seen as a breakthrough that rekindled interest in neural nets and started the movement of "deep learning"[4]. The third event, the discovery of Gene Editing technology, is disclosed in a paper titled "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity," published in *Science* in 2012. Two co-authors of the paper, Jennifer Doudna and Emmanuelle Charpentier, were granted the Nobel Prize in 2020 because of their extraordinary contribution to demonstrating the CRISPR-Cas9 system's potential as a programmable tool for precise gene editing with wide applications in medicine and agriculture.

[4]See https://www.skynettoday.com/overviews/neural-net-history.

2.2. Identification strategy

My identification strategy exploits the unpredictable nature of revolutionary scientific discoveries that have a significant impact on industrial innovation. Firms, especially those that rely on the development of scientific technology, can benefit from the emergence of new technology, which can increase productivity or spur the innovation of new products. Even though firms may have some awareness of ongoing scientific developments, predicting the exact timing of a scientific revolution is still challenging. Therefore, I consider the three academic discoveries—Human Genome Project, Deep Learning and Neural Networks, and Gene Editing and CRISPR technology—as three quasi-exogenous events for existing publicly listed corporations.

Given that there are three events occurring in different years, I first use a stacked difference-in-difference design following Gormley and Matsa (2011), comparing treated firms that operate in areas closely related to the forthcoming scientific breakthrough with control firms that do not operate in areas related to the scientific breakthrough before the breakthrough year. To reduce the estimation bias caused by noisy control firms, I only choose control firms that have never been classified as treated firms in any of the three events, following the literature (Baker, Larcker, and Wang, 2022; Gormley and Matsa, 2011). After classifying treated firms and control firms, I run the following stacked difference-in-difference regression:

$$Y_{i,c,t} = \beta_0 + \beta_1 \cdot \textit{Exposure to Sci-Breakthrough}_{j,c,t} + \alpha_{i,c} + \theta_{t,c} + \epsilon_{i,j,c,t}, \tag{1}$$

where Y is one of the firm outcome variables of interest for firm i and year t, and *Exposure to Sci-Breakthrough* is an indicator that equals 1 if firm i is classified as a treated firm in event cohort c in year t. I examine the window of five years before the breakthrough year and five years after the breakthrough years, $[-5, 5]$. The firm outcome variables include measures of firm performance, *Operating Performance*, and *Log Sales*, as well as measures of innovation outcomes, the number of patents granted *Patent Count*, the citations received *Citation*, and the patent value measure *Log Patent Value*. To account for fixed differences between firms across events, I incorporate the firm-event fixed effect, denoted as $\alpha_{i,c}$. Year-event fixed effects are also included to control for any secular time trends, denoted as $\theta_{t,c}$. Standard errors at the industry level to address the potential covariance among firm-level variables over time within the same four-digit SIC code.

Next, I examine the second research question: the role played by scientific human capital in the face of scientific breakthroughs. Firms that invest more in basic science research within their core business can benefit more from largely unexpected scientific breakthroughs, as they have the absorptive capability and first-mover advantage in transferring new technology into their product and process innovations. Thus, firms with more related scientific human capital in-house prior to the shock should benefit more from the new scientific breakthrough. I run the following stacked triple difference-in-difference regression to test the hypothesis, including an interaction term between the high scientific human capital indicator *HighSHC* and the *Exposure to Sci-Breakthrough*:

$$\begin{aligned} Y_{i,c,t} = \beta_0 + &+\beta_1 \textit{Exposure to Sci-Breakthrough}_{c,t} \cdot \textit{HighSHC}_{i,c,t-3:t-1} \\ &+ \beta_2 \textit{Exposure to Sci-Breakthrough}_{c,t} + \beta_3 \textit{HighSHC}_{i,c,t-3:t-1} \\ &+ \alpha_{i,c} + \theta_{t,c} + \epsilon_{i,c,t} \end{aligned} \tag{2}$$

Where $Y_{i,c,t}$ includes an array of firm-level outcome variables of interest. $\text{HighSHC}_{i,c,t-3:t-1}$ represents indicating whether firm i belongs to the group with high scientific human capital. I measure scientific human capital *SHC* based on the publication relevance with the forthcoming breakthroughs of employees who were hired by firm i between years $t-3$ and $t-1$. I also use the publication relevance of employees who were hired by firm i between years $t-8$ and $t-6$ as a robustness test. The firm-event fixed effects, denoted as $\alpha_{i,c}$, and the year-event fixed effects, denoted as $\theta_{t,c}$, are included. Standard errors at the industry level to address the potential covariance among firm-level variables over time within the same four-digit SIC code. More details about the *SHC* measure are in section 2.3.1.

2.3. Data and variable construction

Treated and control firms. A firm is classified as a treated firm if its patents with the top 10 highest market value for the $[t-3, t-1]$ period have textual similarity to the scientific breakthroughs, ranking in the top 20% among all firms with patents filed during the time window $[t-3, t-1]$. The control firms are composed of the firms ranking in the bottom 80% in terms of technology similarity to the scientific breakthrough, as well as firms without any patents filed in the three years prior to the scientific breakthrough. I only keep control firms that have never been classified as treated firms in any of the three events to mitigate the estimation bias caused by noisy control firms (Baker, Larcker, and Wang, 2022). To obtain the textual similarity between two texts, I utilize the text embedding feature of the natural language processing tool *Instructor-xl*, which can convert text to a vector. Then I calculate the cosine similarity between the vector embedding of each selected patent abstract and the text describing the content of the scientific breakthrough. The firm-level technology similarity in each event is obtained by taking the average of the patent similarity scores of the selected representative patents for firm i as shown below:

$$\text{Technology Similarity}_{i,c} = \frac{1}{N} \sum_{t \in T} \text{cosine similarity between patent}_{t,i,c} \text{ and scientific breakthrough}_c$$

where $\text{Patent}_{t,i,c}$ refers to patents filed by firm i within the past three years $[t-3, t-1]$ that relate to scientific breakthrough c and rank in the top 10 in market value among all patents filed in the same year. $N = \max[30, \text{total number of patents filed by firm } i \text{ during } [t-3, t-1]]$.

Figure 1 displays the ratio of firms that are classified as treated firms within each industry, where the ratio is greater than 10% in at least one of the events. As indicated in the figure, the most shocked industries in the DNA event are Drugs, Healthcare, Business Services, and Laboratory Equipment. Almost 60% of firms in the Drug industry are classified as treated firms. For the Deep Learning event, the most shocked industries include Healthcare, Business Services, Computer Software, Electronics and Equipment, and Finance. The Agriculture, Healthcare, and Drugs industries are the three most influenced industries in the Gene Editing event. The figure demonstrates that the classification is generally consistent with our conception of the three scientific breakthroughs.

Scientific publication data. The publication data used in this study is downloaded from the website OpenAlex, a non-profit organization that collects and publishes the entire database of research papers, book chapters, and other publications covering the whole world (Priem, Piwowar, and Orr, 2022). I keep the publication types including journal articles and proceeding articles that have at least one author affiliated with a company in the U.S. and were published between 1996 and 2017, as 1996 is the fifth year before the first breakthrough event and 2017 is the fifth year after the last breakthrough event. In order to match firms covered by the OpenAlex database with those covered by the Compustat database, I conduct a name match between the names on the publication affiliations and U.S. publicly listed firms. The affiliation of authors is sometimes not the parent firm name. Also, ownership changes can affect the actual parent firm to which a firm belongs. To increase the accuracy of matching between the two databases, I use the data covering the subsidiary-parent relation and ownership relation shared by Arora, Belenzon, and Sheer (2021). In this way, I obtain a dataset of U.S. firm names with Compustat identifiers, including all the corresponding subsidiaries and acquired firms. Finally, I conduct the name match between the affiliation names from the OpenAlex database and the U.S. public firms (including historical parent firm names and all the subsidiary and acquired firm names) and set up a threshold of 70 to keep likely successful matches. For those with a score of 100, I classify them as successful matches. For those below 100 and above 70, I manually check if they are a successful match.

Finally, I obtain about 1,386 firms that have ever published one paper from 1996 to 2017.

2.3.1. Publication-based measure. To measure scientific human capital prior to each scientific breakthrough, I first construct a dataset that tracks the employment history of publishing employees. An author affiliated with a firm who publishes a paper in a given year is considered an employee of that firm for that year. It is acknowledged that publishing employees may not publish annually. It is assumed that an employee does not leave the firm during years without publication, provided the author does not publish with another firm in the interim. If an author publishes for a different firm, the first year of publication for that new firm is considered the year of departure from the previous firm. Subsequently, I collect the employment records of individuals hired by a firm and gather their historical publications up to the year preceding the breakthrough. Then I calculate the annual publication stock using the total number of scientific papers published by all employees (affiliated with firm i in year t) until year t. The measure of scientific human capital *SHC* is defined as the sum of the annual publication stock within the window $[-3, -1]$ centered on the breakthrough year. This is the main measure I use throughout the analysis. I also construct other scientific human capital measures based on a window outside our sample period and a relevant scientific human capital measure based on the closeness of scientists' expertise to a breakthrough using textual information in the publication abstracts and the abstract of a scientific paper. More details are described in Appendix A.

To evaluate the impactfulness of a paper, I create a measure termed *Impactful Paper*. This measure relies on the bibliometric database OpenAlex, which utilizes a state-of-the-art natural language model to classify each paper into 252 subfields and 4,516 topics. A paper is classified as impactful if it ranks among the top 5% in terms of citations received within five years of publication, relative to papers in the same subfield published in the same year. Additionally, an alternative proxy is constructed using a 10% threshold for impactfulness.

The *Impactful Paper* is used to identify star scientists. A scientist is classified as a *Star*

Scientist if more than 50 percent of the papers they have published before a given year t are categorized as *Impactful Paper*.

Patent data. The patent data utilized in this study is sourced from PatentsView, a comprehensive patent database that offers detailed records for approximately seven million patents. This dataset is extracted from the bulk data files of the United States Patent and Trademark Office (USPTO). The dataset encompasses a range of information including patent application dates, grant dates, inventors, assignees, textual content of patents, citations of prior art, and technology classifications, from 1976 to the present. Disambiguation algorithms are employed to assign unique identifiers to patent inventors and assignees, thereby facilitating the tracking of inventors' activities and employment history over time. The Patent-CRSP matching methodology adheres to the approach outlined in Stoffman, Woeppel, and Yavuz (2022). Patent abstract information is employed to assess firms' exposure to scientific breakthroughs and the extent to which their patents rely on scientific knowledge embedded in these breakthroughs. Additionally, citations of prior art are utilized to construct an annual citation network among patents, which is subsequently used to develop an impactful patent identifier.

2.3.2. Patent quantity and quality measure. The patent quantity measure, *Patent Count*, is calculated as the sum of patents granted to firm i in year t. To assess patent quality, I employ several proxies. The first proxy, *Citations* is the cumulative sum of citations received by all patents filed by firm i prior to year t for that year t. The second proxy, *Log Patent Value*, is the aggregate patent value, as defined by Kogan et al. (2017), for all patents granted in year t. The third proxy, *Impactful Patent*, is the number of impactful patents granted to firm i in year t. An impactful patent is defined as one that ranks in the top 5% in terms of forward citations received within five years of being granted, relative to patents filed in the same year.

2.3.3. Patent reliance on science measure. To evaluate the relationship between patents and scientific knowledge, I use two variables: *Reliance on Science* and *First to Cite Science*. Leveraging the capabilities of generative AI in text summarization and classification, models such as ChatGPT have increasingly been employed in the literature to extract textual information from corporate disclosures (Jha et al., 2024; Kim, Muhn, and Nikolaev, 2024). The measure *Reliance on Science* is designed to assess the degree to which patents depend on scientific knowledge. For example, in the context of Deep Learning, I utilize the state-of-the-art generative AI model GPT-3.5-turbo developed by OpenAI. For each patent abstract and its assignee firm's name, the following prompt is used:

"Based on the given patent abstract and its assignee firm's name, please answer the following questions successively. Q1. Does the patent rely on the knowledge of machine learning? Answer choices: Heavily, Mildly, NA. Q2. How important is the patent for the products of the assignee firm? Answer choices: Super, Mild, Little. Q3. Briefly explain your choice to Q1 and Q2 in less than 50 words."

Responses are manually reviewed to ensure reliability, and the prompt is adjusted accordingly. For each patent, responses to Q1 are scored as 1 (Heavily), 0.5 (Mildly), or 0 (NA), and responses to Q2 are scored as 1 (Super), 0.5 (Mild), or 0 (Little). The final score for each patent is calculated by multiplying the score from Q1 by the score from Q2. The firm-level measure *Reliance on Science* is obtained by summing the final scores for all patents within the firm for each year. Detailed information about the prompt design is provided in the appendix.

Another measure of patent reliance on science, *First to Cite Science*, is based on whether a patent is among the first to cite related papers. A paper is considered related if it is referenced in the event paper 's list of citations and was published after 1996, or if it directly cites the event paper. A patent is classified as among the first to cite a related scientific paper if it cites the paper within the first three years of its publication. The patent-to-paper citation data is sourced from Marx and Fuegi (2020, 2022)[5]. The firm-level measure is calculated as the count of patents classified as among the first to cite a related scientific paper.

2.3.4. Other Firm-Level Variables. Firm outcome and control variables are obtained from Compustat. The firm performance outcome variables examined include *Operating Profitability*, *Log Sales*, and *Market Value*. *Operating Profitability* is calculated by dividing operating income before depreciation (Compustat item *oibdp*) by total assets (*at*). *Market Value* is derived by subtracting the book value of common equity (*ceq*) from total assets (*at*) and adding the market value of common equity, which is computed as the product of the closing price ($prcc_c$) and the number of shares outstanding (*csho*). To address the potential concern that larger firms may invest more in basic research, which could confound the effect of the scientific human capital measure, I include a control for the interaction between firm size and a dummy variable indicating treated firms in the post-breakthrough period.

2.4. Summary Statistics

After excluding firms without at least one observation in both the pre-event and post-event periods, the final sample comprises 38,588 observations covering 2,244 firms across the three events. Panel A of Table 1 presents the descriptive statistics for the main variables used in the analysis. Panel B illustrates the distribution of treated and control firms for each event. Across the three events, the ratio of treated firms to control firms is approximately 20%. Panel C displays the proportion of publishing and non-publishing firms within the treated and control groups. Approximately 40% of treated firms publish papers, compared to 26% of control firms. Both groups exhibit a significant proportion of firms that have published at least one paper.

[Insert Table 1 Here]

[5]See https://relianceonscience.org/patent-to-paper-citations

3. Scientific human capital, scientific breakthrough, and firm performance

An important driver of firm growth is product innovation. Innovative ideas typically rely on either the redeployment of existing knowledge or the exploration of new knowledge. Existing knowledge is public and accessible to all, making it difficult to distinguish a firm's products from those of its competitors. In contrast, new knowledge represents an opportunity for firms to pioneer entirely new areas. Products based on new knowledge have the potential to be more disruptive and revolutionary (Ahmadpoor and Jones, 2017; Krieger, Schnitzer, and Watzinger, 2022). Therefore, the ability to absorb and convert new knowledge into novel products is crucial for firm growth. A key measure of a firm's capacity to absorb new knowledge is the stock of related scientific human capital, particularly in areas with intensive scientific knowledge. This section leverages three significant scientific revolutions of the 21st century—the Human Genome Project, Deep Learning and Neural Networks, and Gene Editing technology—to examine the impact of investment in scientific human capital on firm growth during these breakthrough periods.

3.1. Scientific breakthrough and firm performance

Before examining the role of scientific human capital, I first assess whether these breakthroughs bring positive opportunities to firms operating in affected areas and quantify the overall effect of the three scientific breakthroughs on firm value. I perform the regression analysis using the stacked difference-in-differences (DD) setting as specified in equation (1), with results presented in Table 2. The explanatory variables include *Operating Profitability*, and *Log Sales*. It is anticipated that affected firms' performance will improve due to increased productivity and more impactful innovation resulting from new opportunities associated with new knowledge.

Panel A displays the results for firm performance. Columns (1) and (3) control for firm fixed effects and year fixed effects, while columns (2) and (4) include firm-event and year-event fixed effects. The coefficients on *Exposure to Sci-Breakthrough* indicate that treated firms experience approximately 2.7% increase in *Operating Profitability* following the scientific breakthrough compared to control firms. This magnitude is significant given that the mean *Operating Profitability* is around 2.9%. Furthermore, the difference in *Log Sales* between treated and control firms is more pronounced in the five years following the scientific breakthrough. Specifically, firms operating in areas related to the breakthrough exhibit an 11.4% higher sales growth compared to firms in unrelated areas. These results suggest that firms engaged in fields related to scientific breakthroughs achieve superior innovation outcomes due to the new opportunities provided by scientific advancements.

To further support this conjecture, I test the impact of scientific breakthroughs on firm innovation outcomes using the same stacked DD setting, as shown in Panel B of Table 2. The dependent variables include *Patent Count*, *Citations*, and *Log Patent Value*. Given the truncation and skewed nature of patent and citation count variables, I employ Poisson regression for settings where *Patent Count* and *Citations* are the dependent variables. Columns (1) and (2) in Panel B show a positive and statistically insignificant coefficient for *Exposure to Sci-Breakthrough*, suggesting that treated firms do not have more patents granted in the five years following the scientific breakthrough compared to control firms. However, columns (2) through (6) indicate that the quality of patents for treated firms improves more than for control firms post-breakthrough, with coefficients that are positive and significant at the 5% level across all specifications. These results provide evidence that firms operating in areas related to scientific breakthroughs demonstrate superior innovation quality compared to peer firms in unrelated areas.

One potential concern is that firms in related areas might experience other growth opportunities unrelated to the scientific breakthrough, which could confound the observed results. To address this, I test for pre-trends in the differences between treated and control firms prior to the scientific breakthrough. Specifically, I test the following specifications:

$$
\begin{aligned}
Y_{i,c,t} = \beta_0 &+ \beta_1 \cdot \text{Pre1Treated}_{i,c} + \beta_2 \cdot \text{Pre2Treated}_{i,c} + \beta_3 \cdot \text{Pre3Treated}_{i,c} + \beta_4 \cdot \text{Pre4Treated}_{i,c} \\
&+ \beta_5 \cdot \text{EventYearTreated}_{i,c} + \beta_6 \cdot \text{Post1Treated}_{i,c} + \beta_7 \cdot \text{Post2Treated}_{i,c} \\
&+ \beta_8 \cdot \text{Post3Treated}_{i,c} + \beta_9 \cdot \text{Post4Treated}_{i,c} + \beta_{10} \cdot \text{Post5Treated}_{i,c} \\
&+ \alpha_{i,c} + \theta_{t,c} + \epsilon_{i,c,t},
\end{aligned}
\tag{3}
$$

where $Y_{i,c,t}$ includes an array of firm-level variables of interest. $\text{Pre1Treated}_{i,c}$ to $\text{Pre4Treated}_{i,c}$ are dummy variables that are equal to one for treated firm i of event c for the first to fourth year before a scientific breakthrough occurs and zero otherwise. $\text{EventYearTreated}_{i,c}$ is a dummy variable that is equal to one for treated firm i of event c for the occurrence year of a scientific breakthrough and zero otherwise. $\text{Post1Treated}_{i,c}\text{Post1Treated}_{i,c}$ to $\text{Post5Treated}_{i,c}$ are dummy variables that are equal to one for treated firm i of event c for the first to five year after a scientific breakthrough occurs and zero otherwise.

Figure 2 illustrates the difference in *Operating Profitability* and *Log Sales* between treated firms and control firms over the [-5, 5] window surrounding a scientific breakthrough event. Subfigure (a) indicates that, in the years prior to the event, the differences in *Operating Profitability* between the two groups remain stable and statistically insignificant, suggesting the absence of clear pre-trends. However, subsequent to the scientific breakthrough, the differences between treated and control firms become positive and statistically significant. Subfigure (b) demonstrates that, in the three years preceding the event, the differences in *Log Sales* between the two groups remain stable. Following the scientific breakthrough, however, the differences between treated and control firms increase substantially and are statistically significant.

The results presented in Table 2 and depicted in Figure 2 support the hypothesis that firms operating in areas related to scientific breakthroughs achieve higher profits and better innovation outcomes compared to control firms when these breakthroughs occur. This evidence suggests that scientific breakthroughs act as positive shocks to firms operating within overlapping technological fields.

[Insert Table 2 Here]

[Insert Figure 2 Here]

3.2. Does Scientific Human Capital play a role?

Next, I examine the role of scientific human capital in facilitating knowledge transfer and the creation of high-quality innovations in response to scientific breakthroughs. Scientific breakthroughs often introduce cutting-edge knowledge that is challenging to absorb and incorporate into new products. Engagement in basic scientific research enhances a firm's absorptive capacity, enabling it to better integrate new knowledge into its innovation processes (Rosenberg, 2010; Cockburn and Henderson, 1998). Consequently, I predict that firms with a higher stock of related scientific human capital will benefit more from scientific breakthroughs over the five years following the breakthrough compared to firms with a lower stock of scientific human capital. To test this prediction, I include an interaction term between *Exposure to Sci-Breakthrough* and the scientific human capital measure *SHC* in a stacked triple difference-in-differences (DDD) regression, as outlined in equation (2). Given that the scientific human capital measure may be highly correlated with firm size—since larger firms are generally believed to have more resources and a greater ability to diversify the risks associated with investment in basic science research —I control for the potential confounding effect of firm size by including an interaction term between firm size in the year prior to the breakthrough and the *Exposure to Sci-Breakthrough* variable.

Table 3 presents the results on firm performance, with dependent variables including *Operating Profitability*, *Log Sales*, and *Log Market Value*. The scientific human capital measure (*SHC*) takes the value of one if a firm ranks in the top 10% among all firms in a scientific breakthrough cohort, and zero otherwise. In columns (1), (3), and (5), I control for firm fixed effects and year fixed effects to account for unobserved firm characteristics and time-series trends in firm outcomes. In columns (2), (4), and (6), I include event-firm and event-year fixed effects to control for unobserved firm characteristics specific to each event and any secular time trends. The coefficients remain stable across different fixed effect specifications. The interaction between *SHC* and *Exposure to Sci-Breakthrough* is positive and significant at the 5% level in almost all columns, except for *Sales* in column (4). The magnitude of the effect is substantial; for instance, treated firms with high *SHC* experience a 32% increase in market value in the five years following the scientific breakthrough, compared to their peers. The results hold with alternative scientific human capital measures, which are reported in Appendix B. These results indicate that treated firms with a greater stock of scientific human capital exhibit significantly higher growth and profitability over the five years following the scientific breakthrough, relative to both treated firms with lower scientific human capital and all control firms.

[Insert Table 3 Here]

The pre-trend of the effect of scientific human capital. To further demonstrate the argument that scientific breakthroughs provide unique opportunities for corporate scientists to leverage their knowledge and integrate it into the innovation process, it is necessary to address the hypothesis of a pre-trend in the difference in operating profitability and market valuation between treated firms with high scientific human capital and their peers. Specifically, if the difference in profitability between treated firms with high *SHC* and their peers does not exhibit a distinct upward jump immediately following the breakthrough year, it would undermine the argument that scientific human capital plays a critical role in driving firm growth post-breakthrough. This is because the role of corporate scientists, who serve as a bridge between cutting-edge scientific knowledge and industry innovation, should be most evident when new scientific knowledge becomes available for exploitation.

To test this, I plot the average difference in market value between treated firms with high *SHC* and those with low SHC over the [-5,5] window around the three scientific breakthroughs. This is illustrated in subfigures (a) and (b) in Figure 3, along with 95% confidence intervals.

As illustrated in subfigure (a), there is no pre-trend in market value between treated firms with high *SHC* and those with low *SHC*, as the coefficients for the pre-trend dummies remain stable. A noticeable increase occurs at year zero, the breakthrough year, with positive coefficients for the post-trend dummies. Following the breakthrough, firms with high SHC tend to experience a more substantial increase in market value compared to their peers. This difference persists up to the fifth year. For comparison, subfigure (b) displays the pre-trend pattern for the peer group, showing the difference in market value between control firms with high *SHC* and and low *SHC*. There is no discernible jump around the breakthrough year for this group.

The patterns in Figure 3 reinforce the argument that scientific breakthroughs are periods when investments in scientific human capital yield substantial rewards and create significant growth opportunities for firms.

[Insert Figure 3 Here]

4. What role does scientific human capital play?

The previous chapter establishes that firms with more scientific human capital in related areas derive greater benefits from scientific breakthroughs. In this chapter, I explore the role of scientists in treated firms in driving innovation and growth.

4.1. Knowledge spillover channel

The three events I examine are science-intensive and demand advanced knowledge and skills to keep pace with the latest academic findings and translate them into products. Firms with a higher stock of relevant scientific human capital at the time of these breakthroughs are better positioned to rapidly absorb new knowledge and capitalize on first-mover advantages in the development of innovative products.

4.1.1. Accumulation of scientific human capital. A firm that prioritizes basic science encourages its corporate scientists to publish research, which is crucial for the ongoing accumulation of scientific human capital among these scientists. Publications not only reflect the personal reputation of scientists but also serve as a measure of their ability to absorb cutting-edge knowledge. Given that scientific breakthroughs create opportunities to explore new areas, corporate scientists are likely to publish more impactful papers following these breakthroughs, which represents an enhancement in scientific human capital beneficial to the firm. Consequently, I expect that treated firms with higher levels of scientific human capital will produce more impactful papers compared to their peers.

Table 4 presents the results. The variables *Log Impactful Papers 1* and *Log Impactful Papers 2* serve as proxies for the overall publication quality of firms. The coefficients align with our expectations, revealing a positive and significant interaction between *Exposure to Sci-Breakthrough* and *SHC*. Treated firms are observed to publish approximately 10% more impactful papers than their peer firms in the five years following a scientific breakthrough. This supports the notion that firms with a greater stock of scientific human capital demonstrate a more pronounced enhancement in their ability to absorb scientific knowledge compared to peer firms.

[Insert Table 4 Here]

4.1.2. Scientist engagement in patenting activities. Corporate scientists are defined as those scientists affiliated with firms within the sample who have published at least one scholarly paper. When new innovation opportunities emerge from scientific breakthroughs, these scientists represent a critical linkage between academic knowledge and firm-level product innovation. Consequently, it is anticipated that these scientists will exhibit increased engagement in patenting activities following scientific breakthroughs in firms with high *SHC*, compared to firms with low *SHC* and control firms.

To assess scientist engagement in patenting activities, a name-matching process is conducted between author names in the OpenAlex database and inventor names in the USPTO patent database. Initially, a dataset of inventor employment history is constructed using the filing year and the assignee firm information, with the earliest (latest) patent filed by an inventor indicating the first (last) year of employment at the firm. Similarly, a dataset of scientist employment history is compiled using publication years to determine the first and last year of employment.

Subsequently, groups of inventors who file patents in a given year and scientists employed by the same firm in that year are identified within the matched sample used in previous analyses. The fuzzy similarity score between inventor and scientist names in each group is calculated, retaining matches with a score of 80 or above, using the commonly employed name-matching package, *fuzzywuzzy*. Matches with a score of 100 are considered successful by default; for scores below 100, manual verification is conducted to confirm successful matches.

The firm-level measure, the *scientist-inventor Ratio*, is derived by averaging the scientist-inventor ratio across all patents granted to a firm in a given year. Table 5 presents the results where *Scientist-Inventor Ratio* serves as the dependent variable. The hypothesis is tested using a triple-stacked Difference-in-Differences (DID) approach as specified in equation (2). The interaction term between *Exposure to Sci-Breakthrough* and *SHC* consistently shows a positive coefficient across specifications. Firms with higher stock of scientific human capital exhibit a 0.017 higher *Scientist-Inventor Ratio* compared to peer firms. This magnitude is significant given that the mean *Scientist-Inventor Ratio* is approximately 0.014. These findings align with the hypothesis that firms with greater scientific human capital experience higher engagement of scientists in patenting activities, providing direct evidence of the knowledge spillover channel.

[Insert Table 5 Here]

4.1.3. Patent Quantity and Quality. If scientists serve as intermediaries between academic knowledge and industrial innovation, one would expect to observe enhanced innovation outcomes for firms possessing greater scientific human capital. To assess patent outcomes, multiple metrics are employed as detailed in section 3. An additional variable utilized in this analysis is the number of impactful patents granted in a given year, labeled *Impactful Patent*.

Table 6 presents the results. Given the skewed and truncated nature of patent count and citation measures, Poisson regression is employed for the analysis of these variables. Across all specifications in Table 6, the coefficient of the interaction term between *Exposure to Sci-Breakthrough* and *SHC* is positive and statistically significant, even after accounting for various fixed effects and control variables. The magnitude of the effect is substantial. For instance, firms with high scientific human capital that are exposed to scientific breakthroughs exhibit a 27% higher patent value over the subsequent five years compared to peer firms.

To test the pre-trend of the effect on patent quality, I plot the average difference in *Impactful Patent* between treated firms with high *SHC* and those with low SHC over the [-5,5] window around the three scientific breakthroughs. This is illustrated in subfigures (c) and (d) in Figure 3, along with 95% confidence intervals.

As shown in subfigure (c), there is no pre-trend in the number of *Impactful Patent* granted between treated firms with high *SHC* and those with low *SHC*, as the coefficients for the pre-trend dummies remain stable. A noticeable increase occurs at year zero, the breakthrough year, with positive coefficients for the post-trend dummies. Following the breakthrough, firms with high *SHC* tend to experience a more substantial increase in market value compared to their peers. This difference persists up to the fifth year. For comparison, subfigure (b) displays the pre-trend pattern for the peer group, showing the difference in *Impactful Patent* count between control firms with high *SHC* and low *SHC*. There is no discernible jump around the breakthrough year for this group.

These findings provide further evidence that enhanced innovation outcomes can be a significant driver of firm growth for those firms with substantial scientific human capital.

[Insert Table 6 Here]

4.1.4. Patent Reliance on Science. A direct approach to evaluate whether firms with greater SHC excel in integrating new knowledge from scientific breakthroughs into their product innovation is to assess the level of reliance on science in their patents. Additionally, firms with substantial SHC are expected to gain a first-mover advantage in absorbing new knowledge. Therefore, I examine the following hypotheses: 1) Firms with higher *SHC* are more likely to have a greater number of patents that are among the first to cite relevant scientific papers. 2) Firms with higher *SHC* are likely to produce patents that demonstrate a greater reliance on the scientific knowledge derived from scientific breakthroughs.

The variable *First to Cite Science* measures a firm's capacity to rapidly incorporate scientific knowledge into its innovation processes. In contrast, the variable *Reliance on Science* serves as a textual-based proxy that quantifies the degree of relatedness between a firm's patent portfolio in a given year and the scientific knowledge associated with a specific breakthrough event. Detailed definitions of these variables are provided in section 2.3.3.

Table 7 presents the results pertaining to these hypotheses. The first two columns offer evidence supporting the knowledge spillover channel, as firms with greater SHC are more likely to be among the first to cite relevant scientific literature in their patents. Columns (3) and (4) provide weaker evidence for this argument. Overall, the results in Table 7 contribute further evidence supporting the knowledge spillover channel.

[Insert Table 7 Here]

4.2. Attracting star scientist

Another channel through which investment in scientific human capital can contribute to firm growth is by enhancing the firm's ability to attract distinguished scientists and inventors, owing to its reputation. Firms that cultivate a robust and supportive culture of basic science and academic publishing are likely to gain a significant advantage in recruiting star scientists in fields related to scientific breakthroughs (Ahmed, 2022). Therefore, it is hypothesized that firms with greater investments in scientific human capital prior to a breakthrough are better positioned to attract star scientists.

Table 8 presents results testing this hypothesis. The dependent variables include *Number of Scientists* and *Number of Star Scientists*. Columns (1) and (2) show a positive coefficient for the interaction term between *Exposure to Sci-Breakthrough* and *SHC*, though this result is only weakly significant at the 10% level. Columns (3) and (4) demonstrate that firms with a high stock of scientific human capital are more likely to hire star scientists, as indicated by the positive and statistically significant coefficient for the interaction term. Specifically, firms with greater SHC hire 1.3 more star scientists than their peer firms in the five years following a scientific breakthrough. This effect is notable given that the mean number of star scientists among all firms is 1.11. The results in Table 8 suggest that while treated firms with more SHC do not necessarily hire more scientists overall compared to their peers, they are indeed more successful in attracting star scientists.

[Insert Table 8 Here]

5. Conclusion

In this study, I investigate how scientific human capital enhances the value creation of publicly listed companies in the context of emerging scientific breakthroughs. I utilize three prominent scientific breakthroughs originating from universities in the 21st century—Human Genome Project, Deep Learning and Neural Networks, and Gene Editing—as exogenous shocks to firms operating in related fields. This provides a unique setting to examine the role of scientific human capital in driving firm growth, as these breakthrough events serve as unexpected shocks that offer new sources of scientific knowledge requiring academic expertise for absorption and integration into the innovation process.

I argue and demonstrate that firms experiencing these scientific breakthroughs and possessing higher levels of scientific human capital exhibit greater improvements in operating performance and market valuation compared to peer firms with lower levels of scientific human capital or those unaffected by the breakthroughs.

Further analysis into the role of scientific human capital reinforces the knowledge spillover channel, wherein corporate scientists absorb and incorporate new knowledge from scientific breakthroughs into the innovation process. I find that firms with high scientific human capital are more likely to publish impactful papers following a breakthrough compared to their peers. This finding supports the hypothesis that firms with substantial scientific human capital are better equipped to absorb and utilize scientific knowledge from basic science in the five years following new scientific breakthroughs. My findings regarding innovation output also align with the knowledge spillover channel. Specifically, for firms with higher stocks of scientific human capital, I observe: (1) increased engagement of corporate scientists in patenting activities, (2) higher quality of innovation, and (3) a faster incorporation of scientific knowledge into patents, resulting in a greater number of patents based on this knowledge in the five years subsequent to scientific breakthroughs.

Additionally, my results indicate that firms with higher stocks of scientific human capital are more successful in attracting star scientists in the five years following scientific breakthroughs. This finding complements the knowledge spillover channel, as the influx of star scientists further enhances the firm's capacity to integrate cutting-edge scientific knowledge into the innovation process.

My findings underscore the critical role of corporate scientists in advancing firm innovation in response to related scientific breakthroughs. They have significant implications for corporate policy concerning investments in research and development. Given the ongoing trend of specialization between universities, which focus on basic research, and firms, which commercialize this research, the role of corporate scientists is particularly significant. They bridge the gap by translating scientific discoveries from academic institutions into product innovations within the private sector. The evidence presented in this study highlights this bridging role of corporate scientists effectively.

References

Acharya, V., and Z. Xu. 2017. Financial dependence and innovation: The case of public versus private firms. *Journal of Financial Economics* 124:223–43.

Acharya, V. V., R. P. Baghai, and K. V. Subramanian. 2014. Wrongful discharge laws and innovation. *The Review of Financial Studies* 27:301–46.

Aghion, P., and P. Howitt. 1990. A model of growth through creative destruction.

Aghion, P., J. Van Reenen, and L. Zingales. 2013. Innovation and institutional ownership. *American economic review* 103:277–304.

Ahmadpoor, M., and B. F. Jones. 2017. The dual frontier: Patented inventions and prior scientific advance. *Science* 357:583–7.

Ahmed, N. 2022. Scientific labor market and firm-level appropriation strategy in artificial intelligence research. Working Paper, MIT Sloan Working Paper.

Arora, A., S. Belenzon, L. C. Cioaca, L. Sheer, and H. Zhang. 2023. The effect of public science on corporate r&d. Working Paper, National Bureau of Economic Research.

Arora, A., S. Belenzon, K. Kosenko, J. Suh, and Y. Yafeh. 2021. The rise of scientific research in corporate america. Working Paper, National Bureau of Economic Research.

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

Azoulay, P., J. S. Graff Zivin, D. Li, and B. N. Sampat. 2019. Public r&d investments and private-sector patenting: evidence from nih funding rules. *The Review of economic studies* 86:117–52.

Babina, T., A. Fedyk, A. He, and J. Hodson. 2024. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics* 151:103745–.

Babina, T., A. X. He, S. T. Howell, E. R. Perlman, and J. Staudt. 2023. Cutting the innovation engine: how federal funding shocks affect university patenting, entrepreneurship, and publications. *The Quarterly Journal of Economics* 138:895–954.

Baker, A. C., D. F. Larcker, and C. C. Wang. 2022. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics* 144:370–95.

Bena, J., and K. Li. 2014. Corporate innovations and mergers and acquisitions. *The Journal of Finance* 69:1923–60.

Brav, A., W. Jiang, S. Ma, and X. Tian. 2018. How does hedge fund activism reshape corporate innovation? *Journal of Financial Economics* 130:237–64.

Cockburn, I. M., and R. M. Henderson. 1998. Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *The journal of industrial economics* 46:157–82.

Derrien, F., A. Kecskés, and P.-A. Nguyen. 2023. Labor force demographics and corporate innovation. *The Review of Financial Studies* 36:2797–838.

Gormley, T. A., and D. A. Matsa. 2011. Growing out of trouble? corporate responses to liability risk. *The Review of Financial Studies* 24:2781–821.

Jha, M., J. Qian, M. Weber, and B. Yang. 2024. Chatgpt and corporate policies. Working Paper, National Bureau of Economic Research.

Kim, A., M. Muhn, and V. V. Nikolaev. 2024. Bloated disclosures: can chatgpt help investors process information? *Chicago Booth Research Paper* 2023–59.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman. 2017. Technological innovation, resource allocation, and growth. *The quarterly journal of economics* 132:665–712.

Krieger, J. L., M. Schnitzer, and M. Watzinger. 2022. Standing on the shoulders of science. *Strategic Management Journal* .

Li, K., F. Mai, R. Shen, and X. Yan. 2021. Measuring corporate culture using machine learning. *The Review of Financial Studies* 34:3265–315.

Marx, M., and A. Fuegi. 2020. Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal* 41:1572–94.

———. 2022. Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. *Journal of Economics & Management Strategy* 31:369–92.

Myers, K. R., and L. Lanahan. 2022. Estimating spillovers from publicly funded r&d: Evidence from the us department of energy. *American Economic Review* 112:2393–423.

Priem, J., H. Piwowar, and R. Orr. 2022. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833* .

Rosenberg, N. 2010. Why do firms do basic research (with their own money)? In *Studies on science and the innovation process: Selected works of Nathan Rosenberg*, 225–34. World Scientific.

Shen, M. 2021. Skilled labor mobility and firm value: Evidence from green card allocations. *The Review of Financial Studies* 34:4663–700.

Sorenson, O., and L. Fleming. 2004. Science and the diffusion of knowledge. *Research policy* 33:1615–34.

Stoffman, N., M. Woeppel, and M. D. Yavuz. 2022. Small innovators: No risk, no return. *Journal of Accounting and Economics* 74:101492–.

Tian, X., and T. Y. Wang. 2014. Tolerance for failure and corporate innovation. *The Review of Financial Studies* 27:211–55.

Figure 1. Affected sectors in three scientific breakthroughs

This figure presents the most affected sectors for each scientific breakthrough. For each sector (based on the Fama-French 48 industry classification approach), I calculate the fraction of firms classified as treated firms. The figure highlights sectors with a fraction greater than 10% in at least one of the three breakthrough events: the Human Genome Project, Deep Learning, and Gene Editing.

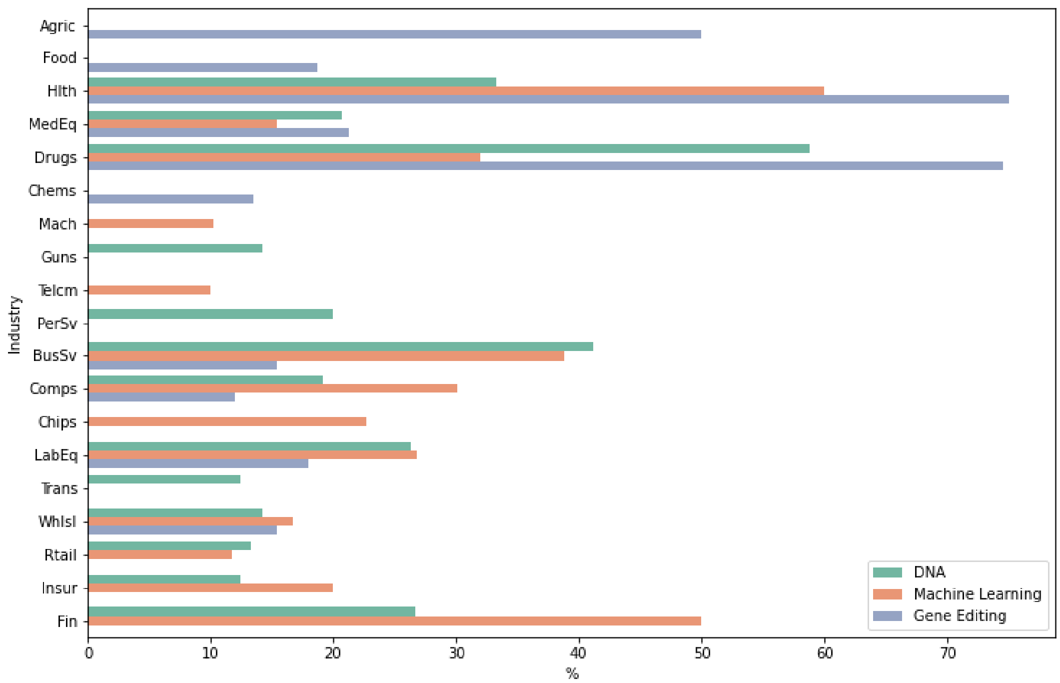
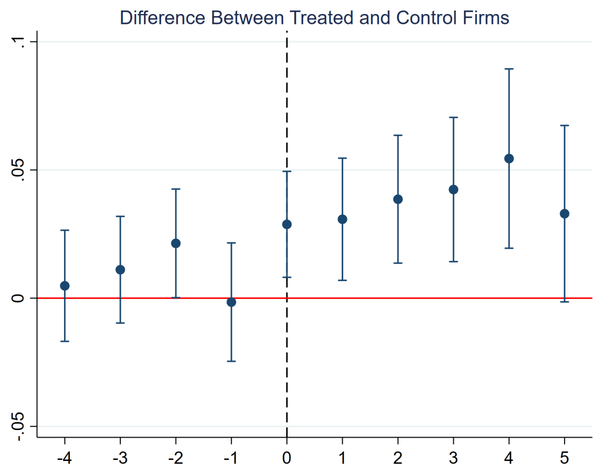
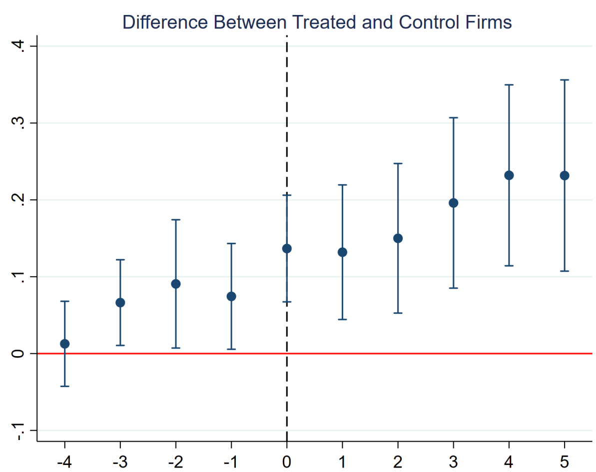


Figure 2. Testing for pre-trends: Profitability and Patent Quality

This figure plots difference-in-difference estimates of the effects of scientific breakthroughs on firm performance, *Operating Profitability* and *Log Sales*. Treated firms are those ranking in the top 20% in terms of technology similarity of firms' core area to a scientific breakthrough; control firms are the remaining 80% and firms without patents filed in the sample period. I include data from an 11-year window centered on the year in which the representative paper of a scientific breakthrough is published. The coefficients are estimated using OLS, as described in Section 3.1. The vertical lines represent 95% confidence intervals. Variables are defined in Appendix A.



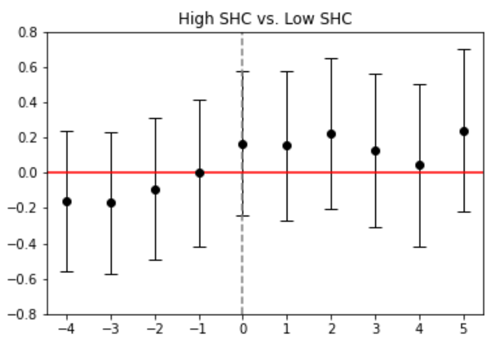
(a) Operating Profitability



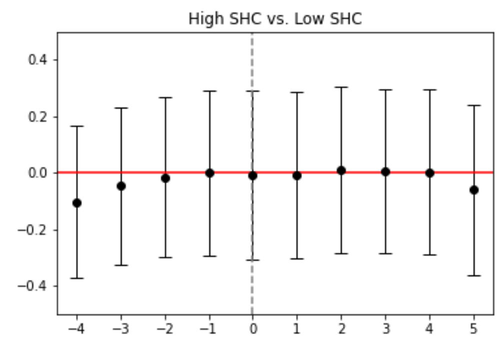
(b) Sales

Figure 3. Testing for pre-trends: High vs. Low Scientific Human Capital

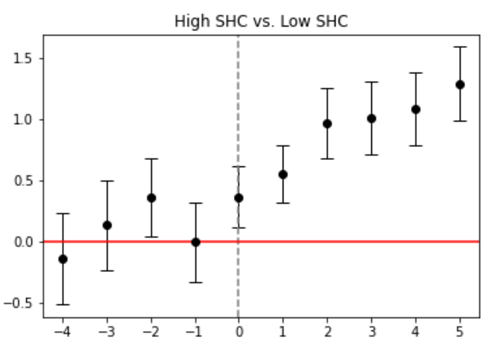
This figure plots estimates of the effects of scientific human capital *HighSHC* on firm valuation *Market Value* and innovation quality *Impactful Patent* within treated firms (subfigures *(a)* and *(c)*) and control firms (subfigures *(b)* and *(d)*), respectively. Treated firms are those ranking in the top 20% in terms of technology similarity of firms' core areas to a scientific breakthrough; control firms are the remaining 80% and firms without patents filed in the sample period. *HighSHC* is a dummy variable that indicates whether a firm possesses high scientific human capital prior to scientific breakthroughs. I include data from an 11-year window centered on the year in which the representative paper of a scientific breakthrough is published. The coefficient on *Market Value* is estimated using OLS, and the coefficient on *Impactful Patent* is estimated using Poisson regression. The vertical lines represent 95% confidence intervals. Variables are defined in Appendix A.



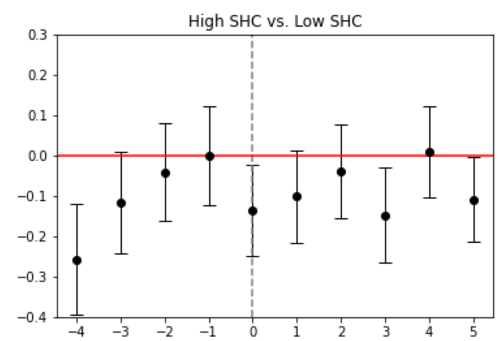
(a) Market Value: Treated Firms



(b) Market Value: Control Firms



(c) Impactful Patent Count: Treated Firms



(d) Impactful Patent Count: Control Firms

Table 1. Summary statistics

Panel A summarizes the descriptive statistics of the key variables used in the analysis. The sample contains 38,588 firm-year observations representing 2,244 firms between 1996 and 2017. Panel B reports the number of treated firms and the number of control firms in each breakthrough event. Treated firms are those ranking in the top 20% in terms of technology similarity of firms' core areas to a scientific breakthrough; control firms are the remaining 80% and firms without patents filed in the sample period. The three scientific breakthrough events include Human Genome Project in 2001, Deep Learning in 2006, and Gene Editing in 2012. Details about scientific breakthrough events are described in section 2.3.1. Panel C reports the number of publishing firms and non-publishing firms in the treated group and control group. Variables are defined in Appendix A.

Panel A: Summary statistics for key variables

	N	Mean	SD	p10	Median	p90
Size ($M)	38,588	5,973.464	21,618.943	27.000	395.000	10,820.000
Operating Profitability	38,588	0.029	0.278	-0.254	0.103	0.221
Sales ($M)	38,588	4,870.901	18,487.477	14.000	424.000	9,545.000
Market Value ($M)	38,588	6.985	2.243	4.142	6.838	10.093
Patent Count	38,588	28.501	109.123	0.000	2.000	47.000
Impactful Patent	38,588	0.405	2.314	0.000	0.000	1.000
Scientist Count	38,588	15.080	72.358	0.000	0.000	19.000
Star Scientist Count	38,588	1.111	7.684	0.000	0.000	1.000
Impactful Paper 1	38,588	3.426	16.974	0.000	0.000	3.000
Impactful Paper 2	38,588	6.165	30.410	0.000	0.000	7.000
Reliance on Science	38,588	0.282	5.578	0.000	0.000	0.000
First to cite Science	38,588	0.060	1.239	0.000	0.000	0.000
Scientist-Inventor Ratio	38,588	0.014	0.071	0.000	0.000	0.000
Scientific Human Capital (SHC)	38,588	0.550	0.931	0.000	0.000	2.269
HighSHC	38,588	0.118	0.323	0.000	0.000	1.000

(continued)

Panel B: The number of treated firms and control firms in each event year

Event Year	Scientific Breakthrough	No. of treated firms	No. of control firms
2001	Human Genome Project	273	1207
2006	Deep Learning and Neural Networks	242	1059
2012	Gene Editing	191	866

Panel C: The number of publishing firms and non-publishing firms in each event year

	Publish		Do not Publish		Total	
	Count	%	Count	%	Count	%
Treated	285	40%	421	60%	706	100%
Control	823	26%	2309	74%	3132	100%

Table 2. Scientific Breakthrough and Firm Performance

This table reports the results from the regression that examines the firm performance after scientific breakthroughs in a stacked difference-in-difference framework. The classification of treated and control firms is described in section 2.3. *Exposure to SciBreak* is a dummy variable that is equal to one for treated firm-year observations in the 5 years after a scientific breakthrough and zero otherwise. The dependent variables in panel A include *Operating Performance* and *Log Sales*. Panel B reports the results of innovation outcome, including *Patent Count*, *Citations*, and *Log Patent Value*. In columns (1), (3) and (5), I control for the firm fixed effects and year fixed effects, In columns (2), (4) and (6), I control for firm × event and year × fixed effects. Standard errors are clustered at the industry level. t-statistics are presented in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Variables are defined in Appendix A.

Panel A: Scientific breakthrough and firm performance

	(1)	(2)	(3)	(4)
	Operating Profitability		Log Sales	
Exposure to SciBreak	0.0273***	0.0277***	0.114***	0.124***
	(3.58)	(3.88)	(3.86)	(3.53)
Firm × Event FEs	No	Yes	No	Yes
Year × Event FEs	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No
Year FE	Yes	No	Yes	No
R-squared	0.699	0.734	0.952	0.965
Observations	38,588	38,588	38,588	38,588

Panel B: Scientific breakthrough and innovation outcome

	(1)	(2)	(3)	(4)	(5)	(6)
	Patent Count		Citations		Log Patent Value	
Exposure to SciBreak	0.168	0.170	0.275**	0.278**	0.170**	0.181**
	(1.44)	(1.42)	(2.45)	(2.39)	(2.48)	(2.48)
Firm × Event FEs	No	Yes	No	Yes	No	Yes
Year × Event FEs	No	Yes	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No	Yes	No
Year FE	Yes	No	Yes	No	Yes	No
R-squared					0.837	0.868
Observations	38,419	38,167	37,993	37,701	38,588	38,588

Table 3. Triple-difference Estimates of the Effects on Firm Performance: The Role of SHC

This table reports the results from the regression that examines the firm operating performance after scientific breakthroughs in a stacked triple difference-in-difference framework. The classification of treated and control firms is described in section 2.3. *Exposure to SciBreak* × *HighSHC* is a dummy variable that is equal to one for treated firm-year observations with high scientific human capital *SHC* in the 5 years after a scientific breakthrough and zero otherwise. The dependent variables include *Operating Performance*, *Log Sales* and *Market Value*. In columns (1), (3) and (5), I control for the firm fixed effects and year fixed effects, In columns (2), (4) and (6), I control for firm × event and year × fixed effects. Standard errors are clustered at the industry level. t-statistics are presented in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Variables are defined in Appendix A.

	(1)	(2)	(3)	(4)	(5)	(6)
	Operating Profitability		Log Sales		Market Value	
Exposure to SciBreak × *HighSHC*	0.0536***	0.0586***	0.172***	0.173**	0.305***	0.279***
	(3.43)	(3.38)	(2.68)	(1.97)	(4.40)	(3.54)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm × Event FEs	No	Yes	No	Yes	No	Yes
Year × Event FEs	No	Yes	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No	Yes	No
Year FE	Yes	No	Yes	No	Yes	No
R-squared	0.702	0.734	0.954	0.965	0.940	0.952
Observations	38588	38588	38588	38588	38588	38588

Table 4. Triple-difference Estimates of the Effects on Corporate Publication

This table reports the results from the regression that examines the firm publication after scientific breakthroughs in a stacked triple difference-in-difference framework. The classification of treated and control firms is described in section 2.3. *Exposure to SciBreak* × *HighSHC* is a dummy variable that is equal to one for treated firm-year observations with high scientific human capital *SHC* in the 5 years after a scientific breakthrough and zero otherwise. The dependent variables include *Log Impactful Paper 1* and *Log Impactful Paper 2*. In columns (1) and (3), I control for the firm fixed effects and year fixed effects, in Column (2) and (4), I control for firm × event and year × fixed effects. Standard errors are clustered at the industry level. t-statistics are presented in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Variables are defined in Appendix A.

	(1)	(2)	(3)	(4)
	Log Impactful Paper 1		Log Impactful Paper 2	
Exposure to SciBreak × *HighSHC*	0.0897**	0.103**	0.0806*	0.0953**
	(2.12)	(2.55)	(1.90)	(2.34)
Controls	Yes	Yes	Yes	Yes
Firm × Event FEs	No	Yes	No	Yes
Year × Event FEs	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No
Year FE	Yes	No	Yes	No
R-squared	0.957	0.979	0.951	0.976
Observations	38,588	38,588	38,588	38,588

Table 5. Triple-difference Estimates of the Effects on Scientist Engagement in Patenting

This table reports the results from the regression that examines scientist engagement in patenting after scientific breakthroughs in a stacked triple difference-in-difference framework. The classification of treated and control firms is described in section 2.3. *Exposure to SciBreak* × *HighSHC* is a dummy variable that is equal to one for treated firm-year observations with high scientific human capital *SHC* in the 5 years after a scientific breakthrough and zero otherwise. The dependent variable is *Scientist-Inventor Ratio*. In column (1), I control for the firm fixed effects and year fixed effects, In column (2), I control for firm × event and year × fixed effects. Standard errors are clustered at the industry level. t-statistics are presented in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Variables are defined in Appendix A.

	(1)	(2)
	Scientist-Inventor Ratio	
Exposure to SciBreak × *HighSHC*	0.0173***	0.0185***
	(2.93)	(3.23)
Controls	Yes	Yes
Firm × Event FEs	No	Yes
Year × Event FEs	No	Yes
Firm FE	Yes	No
Year FE	Yes	No
R-squared	0.529	0.557
Observations	38,588	38,588

Table 6. Triple-difference Estimates of the Effects on Firm Innovation

This table reports the results from the regression that examines the firm innovation performance after scientific breakthroughs in a stacked triple difference-in-difference framework. The classification of treated and control firms is described in section 2.3. *Exposure to SciBreak* × *HighSHC* is a dummy variable that is equal to one for treated firm-year observations with high scientific human capital *SHC* in the 5 years after a scientific breakthrough and zero otherwise. The dependent variables include *Patent Count*, *Log Patent Value*, *Impactful Patent* and *Citation*. In columns (1), (3), (5), and (7), I control for the firm fixed effects and year fixed effects, In columns (2), (4), (6) and (8), I control for firm × event and year × fixed effects. Standard errors are clustered at the industry level. t-statistics are presented in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Variables are defined in Appendix A.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Patent Count		Log Patent Value		Impactful Patent		Citations	
Exposure to SciBreak × *HighSHC*	0.401**	0.386**	0.252**	0.246**	0.755**	0.673**	0.534***	0.531***
	(2.00)	(2.03)	(2.22)	(1.98)	(2.50)	(2.26)	(2.97)	(2.93)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm × Event FEs	No	Yes	No	Yes	No	Yes	No	Yes
Year × Event FEs	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No	Yes	No	Yes	No
Year FE	Yes	No	Yes	No	Yes	No	Yes	No
R-squared			0.839	0.868				
Observations	38419	38167	38588	38588	15680	12834	37993	37701

Table 7. Triple-difference Estimates of the Effects on Patent Reliance on Science

This table reports the results from the regression that examines the level of reliance on science in firm patents after scientific breakthroughs, using a stacked triple difference-in-difference framework. The classification of treated and control firms is described in section 2.3. *Exposure to SciBreak* × *HighSHC* is a dummy variable that is equal to one for treated firm-year observations with high scientific human capital *SHC* in the 5 years after a scientific breakthrough and zero otherwise. The dependent variables include *First to Cite Science* and *Reliance on Science*. In columns (1) and (3), I control for the firm fixed effects and year fixed effects, in Column (2) and (4), I control for firm × event and year × fixed effects. Standard errors are clustered at the industry level. t-statistics are presented in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Variables are defined in Appendix A.

	(1)	(2)	(3)	(4)
	First to cite Science		Reliance on Science	
Exposure to SciBreak × *HighSHC*	0.0362**	0.0484***	0.304*	0.284
	(2.01)	(2.64)	(1.73)	(1.52)
Controls	Yes	Yes	Yes	Yes
Firm × Event FEs	No	Yes	No	Yes
Year × Event FEs	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No
Year FE	Yes	No	Yes	No
R-squared	0.435	0.621	0.435	0.779
Observations	38,588	38,588	38,588	38,588

Table 8. Triple-difference Estimates of the Effects on New Scientist Hiring

This table reports the results from the regression that examines the recruitment of scientists after scientific breakthroughs, using a stacked triple difference-in-difference framework. The classification of treated and control firms is described in section 2.3. *Exposure to SciBreak* × *HighSHC* is a dummy variable that is equal to one for treated firm-year observations with high scientific human capital *SHC* in the 5 years after a scientific breakthrough and zero otherwise. The dependent variables include *Number of Scientists* and *Number of Star Scientists*. In columns (1) and (3), I control for the firm fixed effects and year fixed effects, in Column (2) and (4), I control for firm × event and year × fixed effects. Standard errors are clustered at the industry level. t-statistics are presented in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Variables are defined in Appendix A.

	(1)	(2)	(3)	(4)
	Scientist Count		Star Scientist Count	
Exposure to SciBreak × *HighSHC*	13.00*	13.21*	1.381**	1.395**
	(1.87)	(1.83)	(2.09)	(2.05)
Controls	Yes	Yes	Yes	Yes
Firm × Event FEs	No	Yes	No	Yes
Year × Event FEs	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No
Year FE	Yes	No	Yes	No
R-squared	0.917	0.944	0.901	0.928
Observations	38,588	38,588	38,588	38,588

Appendix A: Definitions of Variables

Variable	Definition
First to cite Science	A patent is considered among the first to cite a related scientific paper if it references the scientific paper within three years of its publication date. A paper is classified as a related paper if it is cited in the event paper published after 1990 or if it cites an event paper. The variable *First to Cite Science* represents the total number of patents granted to firm i in year t that are among the first to cite scientific papers.
Sales	Compustat item *Sales* for firm i in year t..
Market Value	Equal to the total assets (at), subtract the book value of common equity (ceq) and add the market value of common equity (calculated as $prcc_c$ times $csho$).
HighSHC	A dummy variable that equals 1 if the firm ranks in the top 10% in terms of scientific human capital measure *SHC* in a scientific breakthrough event, and 0 otherwise.
HighSHC_Far	A dummy variable that equals 1 if the firm ranks in the top 10% in terms of scientific human capital measure *SHC_Far* in a scientific breakthrough event and 0 otherwise.
HighSHC_Relevant	A dummy variable that equals 1 if the firm ranks in the top 10% in terms of the scientific human capital measure *SHC_Relevant*, and 0 otherwise.
Impactful Paper 1	A paper is classified as an impactful paper if it ranks among the top 10% in terms of citations received within five years of publication, relative to other papers in the same subfield published in the same year. The subfield classification is determined according to the bibliometric database OpenAlex.
Impactful Paper 2	A paper is classified as an impactful paper if it ranks among the top 5% in terms of citations received within five years of publication, relative to other papers in the same subfield published in the same year. The subfield classification is determined according to the bibliometric database OpenAlex.
Impactful Patent	The number of impactful patents granted to firm i in year t. A patent is claassified as impactful patent if it ranks in the top 5% in terms of forward citations received within five years of being granted, relative to other patents filed in the same year.
Scientist Count	An employee is identified as a scientist of firm i in year t if the individual has published papers affiliated with firm i in year t, or if the individual has published papers both before and after year t. *Scientist Count* is the sum of all scientists in firm i in year t.

(continued)

Variable	**Definition**
Star Scientist Count	A scientist is classified as a star scientist if more than 50% of the papers they have published before year t are classified as *Impactful Paper 1*.
Operating Profitability	Divide the operating income before depreciation (Compustat item *oibdp*) by total assets (*at*).
Patent Count	The number of patents granted to firm i in year t.
Reliance on Science	A textual-based measure of a patent reliance on science as described in section 4.1.4.
SHC	The general scientific human capital measure. It is constructed as follows: 1. Obtain the annual publication stock using the total number of scientific papers published by all employees (affiliated with firm i in year t) until year t. 2. The *SHC* is obtained by taking the sum of the annual publication stock within the window $[-3, -1]$ centered on the breakthrough year.
SHC_Far	*SHC_Far* is obtained by taking the sum of the annual publication stock within the window $[-8, -6]$ centered on the breakthrough year.
SHC_Relevant	The relevant scientific human capital measure. It is constructed as follows: 1. Obtain expertise similarity to a breakthrough at the employee level using the maximum similarity score between the abstracts of an employee's historical publication and the abstract of the breakthrough paper. 2. The annual publication relevance score is obtained by taking the sum of the expertise similarity among all employees (affiliated with firm i in year t). 3. *SHC_Relevant* is computed by taking the sum of the annual publication relevance score within the window $[-3, -1]$ centered on the breakthrough year.
Scientist-Inventor Ratio	The average scientist-inventor ratio among all patents granted to a firm in a year. The scientist-inventors are identified by conducting the name match between inventor names and author names for the same firm in a year, as described in section 4.1.3.
Size	Total book assets (Compustat item *at*) of firm i in year t.

Appendix B: Additional Results

Table B.1. Scientific Human Capital and Firm Performance: Alternative SHC measures

This table reports the results from the regression that examines the firm operating performance after scientific breakthroughs in a stacked triple difference-in-difference framework. *HighSHC_Far* is a dummy variable that indicates whether a firm is endowed with high scientific human capital measured in a window $[-8, -6]$ relative to the event year. and zero otherwise. *HighSHC_Relevant* is a dummy variable that indicates whether a firm is endowed with high relevant scientific human capital measured in a window $[-3, -1]$ relative to the event year. and zero otherwise. The dependent variables include *Operating Performance*, *Log Sales* and *Market Value*. In columns (1), (3) and (5), I control for the firm fixed effects and year fixed effects, In columns (2), (4) and (6), I control for firm $\times$ event and year $\times$ fixed effects. Standard errors are clustered at the industry level. t-statistics are presented in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Variables are defined in Appendix A.

Panel A: Scientific human capital based on early years

	(1)	(2)	(3)	(4)	(5)	(6)
	Operating Profitability		Log Sales		Market Value	
Exposure to SciBreak $\times$	0.0530**	0.0526**	0.316***	0.262***	0.321***	0.270***
HighSHC_Far	(2.41)	(2.07)	(4.13)	(3.22)	(4.52)	(3.76)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm $\times$ Event FEs	No	Yes	No	Yes	No	Yes
Year $\times$ Event FEs	No	Yes	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No	Yes	No
Year FE	Yes	No	Yes	No	Yes	No
R-squared	0.702	0.734	0.954	0.965	0.940	0.952
Observations	38,588	38,588	38,588	38,588	38,588	38,588

Panel B: Scientific human capital based on expertise similarity to scientific breakthroughs

	(1)	(2)	(3)	(4)	(5)	(6)
	Operating Profitability		Log Sales		Market Value	
Exposure to SciBreak $\times$	0.0519***	0.0548***	0.173***	0.116*	0.336***	0.284***
HighSHC_Relevant	(3.51)	(3.31)	(2.82)	(1.90)	(4.35)	(3.72)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm $\times$ Event FEs	No	Yes	No	Yes	No	Yes
Year $\times$ Event FEs	No	Yes	No	Yes	No	Yes
Firm FE	Yes	No	Yes	No	Yes	No
Year FE	Yes	No	Yes	No	Yes	No
R-squared	0.702	0.734	0.954	0.965	0.940	0.952
Observations	38,588	38,588	38,588	38,588	38,588	38,588