

Of Mice and Academics: Examining the Effect of Openness on Innovation

Fiona Murray*, Philippe Aghion†, Mathias Dewatripont‡
Julian Kolev§ and Scott Stern¶

November 9, 2010

Abstract

This paper argues that openness of upstream research raises the incentives for additional upstream research by encouraging the establishment of entirely new research directions. We test this hypothesis by examining a “natural experiment” in openness within the academic community: NIH agreements signed during the late 1990s that limited the IP restrictions imposed on academics regarding certain genetically engineered mice. Using a sample of engineered mice that are linked to specific scientific papers (some affected by the NIH agreements and some not), we implement a differences-in-differences estimator to evaluate how the level and type of follow-on research using these mice changes after the NIH-induced increase in openness. We find a significant increase in the level of follow-on research, and, more importantly, we find that openness encourages the exploration of more diverse research paths. Moreover, we find that increased openness does not reduce the number of new mice created. Overall, our findings highlight a neglected cost of IP: reductions in the diversity of experimentation and in the flow of new lines that result from cross-fertilization with existing lines..

1 Introduction

Over the past three decades, there has been a significant increase in the scope of formal intellectual property (IP) rights, such as patents, over scientific knowledge traditionally maintained in the public domain (Mowery, et al, 2001; Murray, 2002; Heller, 2008). This dramatic expansion of IP rights for early-stage research tools has spurred a wide-ranging policy debate, with particular attention paid to the impact that intellectual property can play in shaping incentives

*MIT, Sloan School of Management

†Harvard University and NBER

‡ECARES, Universite Libre de Bruxelles and CEPR

§Harvard University

¶Kellogg School of Management, Northwestern University and NBER

for both fundamental scientific research and downstream innovation (Merges and Nelson, 1990; Gallini and Scotchmer, 2002).

The impact of intellectual property rights on the rate and direction of research is subtle. On the one hand, under what we might call the “appropriability” view, early-stage IP simultaneously encourages the establishment of new research lines but discourages downstream exploitation. When strong and broad IP rights over upstream tools are available, upstream researchers can avoid expropriation by downstream researchers, thus enhancing the incentives for upstream researchers while imposing a tax on downstream research activities (Scotchmer, 1996; 2004). According to the appropriability perspective, openness enhances incentives for downstream exploitation along existing research lines, but also reduces the number of new research lines that are started.

A second approach, which we refer to as the “control view” (see Aghion, Dewatripont and Stein (2009), henceforth ADS), emphasizes the role of intellectual freedom for researchers - defined as the granting of control rights allowing researchers to select their research direction. In the control view, new lines can be initiated by the speculative use of existing lines. By improving the matching between researchers and ideas, openness increases the likelihood that a new idea will be developed by a sequence of researchers until it becomes fully mature and commercializable.

Compared with the control view, the appropriability view abstracts away from two fundamental features of knowledge. First, a single upstream idea can, in principle, be applied across multiple later-stage domains and applications (Bresnahan and Trajtenberg, 1995). In other words, ideas are non-rivalrous. Second, it may be extremely difficult in advance to precisely articulate the diversity and range of applications arising from a given upstream idea. Different individuals may have different perceptions regarding the main application of an idea or the follow-on research projects they would prefer to pursue. In other words, rather than focusing exclusively on the value generated along a single line, we argue that it is important to recognize that multiple researchers may seek to pursue a diverse range of exploratory, horizontal follow-on experiments, each of which may initiate new (potentially unanticipated) research lines.

The control approach suggests three main channels whereby a reduction in the scope of upstream IP rights, or an increase in upstream openness, can influence the level and nature of scientific research. ¹ First, by reducing the costs of accessing key research inputs, openness encourages new researchers to enter, thus increasing the diversity of academic research participants. ² Second,

¹Interestingly, while prior research highlights the potential for gridlock arising from an upstream patent thicket, little attention has been paid to the interaction between the openness of scientific knowledge and the diversity of scientific experimentation across multiple research lines. In this paper we examine the impact of changes in openness on the level and nature of research in a setting in which exploration is particularly salient: academic research. Our analysis builds on a literature exploring the distinctive incentives and control rights provided by the institutional regime for research in academia as compared to industry (Dasgupta and David, 1994; David, 2003; Stern, 2004).

²Note that in our analysis we do not deal with the complementary issue of how to finance and/or reward the discovery of the initial research input. This could happen either through

openness makes researchers with high levels of freedom (academics) more likely to engage in experiments that broaden the horizontal diversity of research lines, in part because subsequent openness implies that their research can itself have subsequent impact across a wide range of research lines. Finally, there is the expropriation effect whereby an increase in the level of openness of upstream research reduces the costs associated with its exploitation along a given vertical research line.

Overall, in research settings characterized by high levels of freedom, openness will do more than simply increase the overall level of research output: it should also be closely associated with more diverse and exploratory research, as well as the establishment and exploration of entirely new research lines. Moreover, while openness should affect both basic and applied research, we predict the impact on basic research will dominate when researchers in the pre-openness period face high fixed costs of initiating a new line of research. By contrast, the increase in applied research will dominate when significant basic research has already been conducted.

We evaluate these empirical implications by taking advantage of a “natural experiment” in openness that occurred in the late 1990s in the field of mouse genetics. The experiment resulted from two Memoranda of Understanding (MoU) between DuPont and the National Institutes of Health (NIH) regarding the ability of academic researchers to gain access to hundreds of genetically engineered mice developed using two types of genetic modification technology: Cre-lox and Onco, both covered by patents owned or licensed by DuPont. Prior to the NIH-MoUs, DuPont adopted stringent restrictions on use of the mice for academic research. However, the MoUs lifted these restrictions by implementing a simple contract, providing a royalty-free and costless license that removed any claims to reach-through rights on downstream research, and ensuring that the mice covered under the patents would be made available through the Jackson Laboratory (the world’s largest non-profit repository for research mice). The NIH-MoUs constitute an openness shock for the mouse genetics research community: Prior to the MoUs research tools covered by the patents - hundreds of varieties of Cre-lox or Onco mice developed in the early 1990s - were subject to stringent restrictions in openness. After the MoUs they suddenly became widely accessible to the entire academic research community.

Our empirical approach takes advantage of key aspects of this natural experiment to develop and implement a differences-in-differences estimate of the impact of the NIH-MoU openness experiment on both the level and nature of follow-on research. First, each genetically engineered mouse is associated with a journal article that describes its initial development; we refer to these publications as “mouse-articles.” We are able to construct a treatment sample based on “mouse-articles” affected by the NIH agreements, and a control sample of mouse-articles unaffected by the agreements (composed of mice created using either Knock-out or Spontaneous genetic modification). Second, both the timing

publicly subsidized research or through public buy-outs of (private) patents as in Kremer (1998).

and the scope of the NIH agreements were unanticipated by the mouse genetics community. As a result, there was an unexpected and dramatic shift in the level of openness in a short period of time. Finally, we are able to take advantage of detailed bibliometric data for articles citing the mouse-articles in either the treatment or control groups to characterize how the openness shock changed the nature of subsequent research (relative to the evolution of citations within the control group).

To implement this empirical approach, we analyze the citations to a sample of more than 2000 mouse-articles, approximately 10% of which described Cre-lox and Onco mice that experienced a shift in the level of openness as the result of the NIH agreements. By comparing citations to the mouse-articles before and after the agreement (and comparing them to the evolution of citations in the control sample), we are able to isolate the causal impact of a shift in scientific openness on the level and nature of follow-on research. In particular, rather than simply examine whether there is a net increase or decrease in the level of citations, the bulk of our analysis examines how the *composition* of citations differs after the openness shock. Specifically, we construct measures capturing whether the research community using a particular mouse in any given year is composed of authors new to the research line (i.e. the number of new authors citing the mouse-article), whether research is associated with the establishment of new research lines that had not previously used a particular mouse (i.e. whether the citations include keywords that had never before been linked to a particular mouse-article), and whether the research is basic and “upstream” or applied and “downstream” in nature (captured by the type of journals in which the citations are published). Finally, we analyze the effects of the two NIH-MoU openness shocks on the flow of new mouse-articles. While the appropriability approach unambiguously predicts that increased openness should lead to a reduction in the number of new mouse-articles, the control approach is consistent with a positive or neutral effect of increased openness on mouse creation.

Our results can be summarized as follows: First, the NIH agreements result in a significant increase in the level of follow-on research. More importantly, the bulk of the new citations arise from articles published by “new” researchers or institutions. In other words, the boost in citations to a given mouse-article in the post-NIH agreement period comes from researchers that had not cited that mouse-article prior to the NIH agreement. Next, our results offer direct evidence that increased scientific openness is associated with the establishment of entirely new research lines. Specifically, the openness agreements lead to a significant increase in the diversity of the journals in which mouse-articles in the treatment group are cited, and, perhaps even more strikingly, a very significant increase in the number of previously unused “keywords” describing the research contributions of the citing articles. Finally, we find that increased openness is *not* associated with a reduction in the number of new mouse-articles; instead, the number of new mice created either increases or remains the same after the openness shocks. These findings survive a broad range of robustness tests, and are consistent with the control view of innovation. In particular, the result that

the number of new mice do not decrease after the openness shocks is at odds with the predictions of the appropriability approach. Under the appropriability story, new research lines can only result from the introduction of new tools (in this case, new mice), and greater openness reduces the incentives for their creation. By contrast, our findings are more consistent with the control view, where new lines can also result from cross-fertilization with existing lines by free researchers, and that this cross-fertilization is itself facilitated by openness.

The paper is organized as follows. Section 2 presents the appropriability and control views of the effects of openness on scientific knowledge production. Section 3 describes the experiment we use to explore the effects of increased openness on the horizontal and vertical aspects of research flows. Section 4 outlines our identification strategy. Section 5 presents the data and summary statistics. Section 6 presents the empirical results. Section 7 concludes.

2 Two views of the role of openness in scientific knowledge production

We broadly define openness as any event or device that increases a researcher's ability to access the ideas or materials of other researchers. Alternatively, it allows a researcher to provide access to her own ideas and share them as she sees fit. There are two main approaches to analyze the effects of openness on basic research and innovation. The appropriability approach stresses the idea that openness should favor more applied research, generally at the expense of more basic research, as it reduces the extent to which upstream researchers can appropriate the returns from their own research.³ The control approach⁴ predicts instead that increased openness should foster basic research and the creation of new lines, in particular by reducing researchers' cost of accessing other researchers' ideas, thereby making it more likely that the alternative strategies pursued by researchers with high levels of freedom will actually lead to new lines. In particular, this latter approach emphasizes the complementarity between openness and freedom and the resulting effect of openness on the diversity of subsequent research lines.

2.1 The appropriability view

Consider a two-stage research line. Each stage requires one researcher. Success at each stage, occurs with probability p , and moves research to the next stage. As long as we focus on a single research line, a first effect of openness is that it increases the extent to which stage 2 can extract rents from stage 1. Thus, if V denotes the ex post value of the line (e.g the price at which the research can be

³Although openness makes it also more likely that stage- $i + 1$ researchers will know about, and therefore build upon, the ideas of stage- i researchers, which in turn should increase the ex ante incentive to undertake stage i .

⁴See Aghion-Dewatripont-Stein (2009).

commercialized), then the value Π_2 of the line as of stage 2, is equal to

$$\Pi_2 = pV + \psi - w,$$

where ψ is the additional rent openness gives stage 2 at the expense of stage 1, and w is the wage paid to a researcher (we take it as given for simplicity). The stage-1 value of the line can then be expressed as:

$$\Pi_1 = p(pV - \psi - w) - w = p^2V - p\psi - (1 + p)w.$$

Thus, trivially, increasing ψ fosters stage-2 research at the expense of stage 1 research since it raises Π_2 and reduces Π_1 . In this framework, openness is expected to reduce the incentives to create new lines (as it increases the probability that Π_1 will be less than the fixed cost K of creating a new line). Meanwhile, openness should foster increased downstream research thanks to higher appropriability.⁵ Finally, note that the appropriability approach has nothing to say on the relationship between openness and the diversity of research lines.

2.2 The control view and the effect of openness on diversity

In this subsection we introduce the notion of academic freedom, and then analyze the interplay between freedom and openness. We argue that to the extent that early research stages are optimally managed under academic freedom, openness in early stages of research should foster the creation of new research lines.

2.2.1 Introducing academic freedom

We maintain the assumption that research proceeds along multi-stage lines, with each line starting with an initial idea I_0 , and eventually generating a marketable product with value V after $k \geq 2$ successful stages. As before, we assume that it is sufficient to hire one researcher per stage, and that this researcher obtains

⁵One can extend the analysis by assuming that openness has an additional effect, namely to increase the possibility for the stage-1 researcher to transmit her research to stage 2 researcher(s). Indeed, once success has been obtained in stage 1, it may not be immediate to identify a researcher who will be able to carry the project forward into stage 2. This may require a ‘successful match’, whose probability will naturally rise with openness. Specifically, we call the probability of such a match A and we assume it depends positively on ψ . This means the stage-1 value of the line becomes:

$$\Pi_1 = pA(\psi)(pV - \psi - w) - w = A(\psi)(p^2V - p\psi) - (1 + p)w.$$

In turn, this implies:

$$\frac{d\Pi_1}{d\psi} = A'(\psi)(p^2V - p\psi) - pA(\psi),$$

which can be positive in particular if the effect of openness on the quality of matching is high (i.e. if $A'(\psi)$ is high). If we assume that research is socially optimal (i.e. if $p(pV - w) - w > 0$), then a sufficient condition for openness to be efficiency-enhancing is that $\frac{d\Pi_1}{d\psi} > 0$.

a probability of success equal to $p < 1$ at any stage if he follows the success-maximizing (“practical”) research strategy at that stage. But now we also assume that, instead of the practical strategy, a researcher is free to follow an “alternative” strategy. If we assume that the scientist has a zero individual probability of success following this approach, then this alternative strategy amounts to the scientist working on an activity that he enjoys more but that does not pay off in monetary terms. However, as we describe at the end of this section, we can interpret this alternative strategy as the case in which the scientist works on an activity that may help initiate new lines but does not generate progress on the initial line.

There is an infinite supply of researchers at each stage, each of whom has an outside option w . After being hired at stage j , the scientist is exposed to idea I_{j-1} , and then learns whether he would prefer following the practical strategy or the alternative strategy. If he is able to undertake his favored strategy, he suffers no disutility from working. If, however, the scientist has to undertake the strategy that he likes less, he suffers disutility of z . The ex ante probability that a scientist prefers to follow the practical strategy is given by α . Assume further that the choice of the practical vs. alternative strategy is ex ante non-contractible.⁶

Academic research (or freedom) differs from private-sector research in that it leaves control rights over the research strategy in the hands of the researcher. Thus if a research line is pursued in academia, the researcher is paid wage w and always works on his preferred strategy. This implies that with probability α , the scientist works on the practical strategy, and with probability $(1 - \alpha)$, he works on the alternative strategy. Thus the ex ante probability of advancing to the next stage is given by αp . Now consider a researcher employed by the private sector. Whether the researcher prefers the practical or the alternative strategy, becomes evident once the researcher has been hired by the firm and has been given access to the idea by the firm owner. Yet ex post, the firm owner has the authority to force the scientist to work on the practical strategy. Anticipating this, the researcher will demand a wage of $w_p = w + (1 - \alpha)z$ in order to work in the private sector. The $(1 - \alpha)z$ markup over the academic wage represents compensation for loss of creative freedom—the fact that scientists now always have to adopt the practical strategy, whether this turns out to coincide with their preferences or not.

2.2.2 Complementarity between openness and freedom: diversification effects

A main finding in ADS is that academic freedom tends to dominate private sector focus at earlier stages on a research line. To see this, take a research line involving 2 stages, and suppose that the first stage has been successful, so that we are now at stage 2, with one more stage to be completed in order to generate

⁶In other words, one cannot write a contract that promises a scientist a bonus for following the practical strategy, because the nature of what kind of work that strategy entails cannot be adequately described ahead of time.

a payoff of V . If this last stage of research is done in the private sector, the expected payoff is equal to $E(\pi_2^p) = pV - w_p$. If instead the last stage is done in academia, the expected payoff is equal to $E(\pi_2^a) = \alpha pV - w$. This means that private sector research will yield a higher payoff than free (academic) research and only if $(1 - \alpha)pV > (w_p - w)$, or equivalently $pV > z$.

Now, let Π_2 denote the maximum of $E(\pi_2^p)$ and $E(\pi_2^a)$. Moving back to stage 1, we now compare between $E(\pi_1^p) = p\Pi_2 - w_p$ and $E(\pi_1^a) = \alpha p\Pi_2 - w$. Private sector research will yield a higher payoff than free (academic) research at stage 1 if and only if $p\Pi_2 > z$.

Since $\Pi_2 < V$, it follows that private sector research is value-maximizing at stage 1, it is also value-maximizing at stage 2. In particular it cannot be value maximizing to have academic freedom operate at later stages than private sector research. The key result is therefore that academic freedom will be the optimal governance structure at earlier stages and private sector research will be optimal at later stages. The intuition is that while academia's wage cost advantage stays constant over research stages, its lower probability of success becomes more problematic as one approaches the final value V .

This result can be generalized to lines of any length k : if Π_i denotes the NPV of the line of length k as of stage i , we have:

$$\Pi_i = \max\{E(\pi_i^p) = p\Pi_{i+1} - w_p, E(\pi_i^a) = \alpha p\Pi_{i+1} - w\} < \Pi_{i+1}.$$

This monotonicity property, together with the fact that research should be pursued under academic freedom if and only if $p\Pi_{i+1} > z$, yields the desired result.⁷

That more openness should foster the creation of new lines, follows from the fact that openness favors the cross-fertilization of ideas within stages. More formally, consider two parallel research lines, 1 and 2, each of which operates as described above. Namely, with ex ante probability α the researcher initially allocated to the current stage of either of these two lines, prefers to pursue the practical strategy for that line whereas with probability $(1 - \alpha)$ he prefers not to pursue this practical strategy. Now openness implies that the scientist on line 1 can learn about project 2 and vice-versa, and that consequently with positive probability φ , thanks to academic freedom and the resulting horizontal interaction, she may choose to work on the practical strategy for project 2 if nobody else does. A greater degree of openness implies a higher value of φ .

Openness also increases the net present value of a research line operated

⁷Note that the model so far provides a rationale for free (academic) research even in the extreme case where the alternative strategy has no value beyond saving the researcher the disutility of pursuing the practical strategy. In reality however there is value in experimenting with ideas that may lead to an entirely new research lines, consistently with the idea that scientific discoveries do not follow a purely "linear" model. This does not alter the relative optimality of academia (vs. private research) in earlier (vs. later) stages of research. It does, however, raise the desirability of freedom in general (and academia as the institutional regime that supports such freedom), if we make the realistic assumption that pursuing the alternative strategy confers a higher probability of generating entirely new research lines than pursuing the practical strategy (note that, realistically, the probability of such an event, possibly the result of an "accidental" discovery, is nonzero for both strategies). See ADS for details.

under academic freedom in a given stage i , from:

$$\alpha p \Pi_i - w$$

to:

$$[\alpha + (1 - \alpha)\varphi]p \Pi_i - w.$$

Thus openness increases the social value of operating any stage (particularly earlier stages) under academic freedom. The idea that openness favors cross-fertilization also implies that it should widen the pool of researchers and research institutions working on a particular research idea, since one key feature of academia is the fact that diverse researchers experiment with scientific ideas to investigate their full potential. What openness does is to reduce the fixed cost of ‘entering’ a particular research area to conduct these investigations.

Remark 1: An additional reason (see ADS) to expect that increased openness should foster free research and therefore the creation of new lines is based in the fact that academic institutions are typically not-for-profit. This helps to enforce the commitment not to monitor any individual scientist’s research agenda, which in turn makes less likely to incur the cost of monitoring researchers. But this same feature also implies that a reduction in the cost of accessing research inputs should make a bigger difference for academic research than for private sector research.

Remark 2: If openness enhances basic research and the creation of new lines, this implies that it should have a long-lasting effect on the flow of subsequent publications. This is because new lines take a significant amount of time before maturing, and their development could lead to even more research lines being created. Indeed, starting a new line means a positive probability of a long dynamic flow of new discoveries whose research lines continue long after the original line has ended.

2.3 Contrasting the predictions from the two approaches

The appropriability approach predicts that an increase in the openness of in basic research should discourage further basic research and reduce the creation of new lines, while encouraging greater levels of downstream research and innovation. By contrast, the control approach generates at least three distinct predictions. First, an openness shock should increase the diversity of researchers engaged in follow-on innovation. With more open and independent access to innovation inputs, new researchers can overcome fixed cost barriers to move from other fields and build on these inputs. Second, an openness shock should increase the flow and diversity in the types of research that are being pursued, as it fosters horizontal experimentation, therefore leading to the creation of new lines. Third, openness should have a different impact on basic or applied research. In particular when controls rights conditions are the first order consideration of the openness shift, then we would anticipate that the vertical exploitation outcome would dominate. However, when access costs are initially high or when control rights considerations are not first order, then we would expect the boost

in openness to affect basic research with horizontal exploration dominating. Finally, under the control view, increased openness need not reduce the rate of creation for new research lines. This is because openness facilitates greater levels of cross-fertilization by free researchers with existing lines, which works to offset the reduced appropriability for early-stage researchers when their ability to restrict access to their IP is weakened.

We now confront these contrasting sets of predictions with the empirical findings from our natural experiment of increased openness in genetically engineered mice.

3 Empirical setting: experiments in the openness of genetically engineered mice

3.1 Towards genetically engineered mice

In this section we describe two “natural experiments” that significantly shifted the level of openness associated with two broad categories of genetically engineered mice - both crucial inputs into cumulative research in the modern life sciences.⁸ To understand how we make use of these shifts in openness, it is useful to consider the essential role played by specialized research mice in modern life sciences research. With their genetic likeness to humans (the mouse and human genomes have a 99% similarity), mice play a central role in the study of cancer and other human diseases. Throughout the twentieth century, scientists in mouse genetics relied on “spontaneous mutations” for their disease studies: researchers bred mice that naturally exhibited particular disease-linked symptoms or behaviors.⁹ To facilitate their efforts, the research community developed open access institutions, notably the Jackson Laboratory (a mouse repository in Bar Harbor, Maine) to classify, breed, and distribute specialized research mice to the academic community.¹⁰ In the early 1980s, advances in molecular biology and the ability to manipulate embryonic stem cells allowed researchers to develop a set of systematic and precise methodologies for engineering specialized mice as research tools, greatly expanding the application of research mice in life sciences research.¹¹ Three breakthroughs were particularly important. First, in a discovery awarded the 2007 Nobel Prize in Medicine, Mario Capecchi of

⁸We refer the reader to Murray (2009) for more details and references on the history of intellectual property and openness in the mouse genetics community.

⁹Given the value of such mutations, researchers also developed techniques to significantly increase the rate of mutation of research mice such as the exposing pregnant mice to high levels of radiation (See Murray (2007,2009)).

¹⁰See Rader (2004).

¹¹The use of these methods for mouse engineering are complex and costly. To create a mouse with particular genes inserted within a mouse genome, scientists must first inject foreign DNA into mouse eggs, transplant the eggs into female mice, and, if successful, monitor and breed the incorporation of the genes into the offspring. During our sample period, the development of a “mouse line” from scratch likely involved at least 18 months of laboratory research and a significant investment of time and attention by a principal investigator (Rader, 2004; Murray, 2009).

the University of Utah and his collaborators developed “Knock-out” technology, enabling researchers to delete specific genes in research mice. Second, with partial funding from DuPont Corporation, Professor Phillip Leder at Harvard University developed Oncomouse methods, which provided a means for inserting (rather than deleting) genes into an embryo, thereby making mice susceptible to particular forms of cancer and other diseases. Finally, researchers in the life sciences division of DuPont developed the Cre-lox technology - a precise “cutting and pasting” tool that turns off genes in specific tissue or organs.

By offering general-purpose tools to engineer discrete changes in the genetic profile of research mice, each of these three methods contributed to a paradigm shift in life sciences research. These tools gave scientists a means to investigate a wide variety of new research problems, from very basic research on the impact of genetic variation on disease incidence to the development and optimization of new therapies.¹²In practical terms they allowed researchers to develop three new types of research mice: Knock-out, Cre-lox and Onco mice to be used as critical research inputs in their experiments in addition to the more traditional spontaneous mice.

The revolution in mouse genetics occurred alongside several important shifts in the role of formal IP in life sciences research. In 1980, the Supreme Court decision in *Diamond v Chakrabarty* established the patentability of genetically engineered organisms and the Bayh-Dole Act affirmatively allowed universities to seek patent protection and licensing revenues from Federally-funded research.¹³ While many observers took universities growing patent portfolio as an indicator of evolving role of universities as engines of innovation and commercialization (Henderson, Jaffe and Trajtenberg, 1998), some argued that strong IP rights over scientific research discoveries were detrimental to research productivity and cumulative discovery (Heller and Eisenberg, 1998). In particular, some universities placed significant restrictions on the distribution of patented research materials to academic researchers (e.g., the University of Wisconsin restricted the open distribution and use of patented stem cell lines while other universities were accused of rent-seeking when they sought to enforce IP claims over independent commercial discoveries (e.g. the University of Rochester’s enforcement of its patents on the Cox-2 pathway¹⁴).

¹²The 2007 Nobel Prize announcement regarding knock-out mice states that “Almost every aspect of mammalian physiology can be studied by gene targeting. We have consequently witnessed an explosion of research activities applying the technology. Gene targeting has now been used by so many research groups and in so many contexts that it is impossible to make a brief summary of the results.” (Nobel Prize Press Release http://nobelprize.org/nobel_prizes/medicine/laureates/2007/press.html).

¹³These legal and policy shifts reflected, in part, increasing appreciation that certain types of academic research were increasingly dual in nature: fundamental scientific discoveries that could simultaneously have a high degree of commercial utility (Murray and Stern, 2007)

¹⁴See Murray (2007).

3.2 Research under limited openness

Debates over the role of patents on scientific research tools were particularly salient for researchers exploiting the transformation in mouse genetics. All three of the key mouse engineering tools and the mice generated with them – Knock-out mice, Onco mice and Cre-lox mice – were covered by relatively broad patents.¹⁵In the case of Knock-out mice, the University of Utah received a patent in 1987 but never sought to enforce the patent against follow-on researchers using the Knock-out methodology. Instead, Knock-out mice were made available at (essentially) marginal cost through the Jackson Laboratory. The patents over the Onco and Cre-lox technologies proved to be much more controversial. As a result of their partial funding of Harvard’s Oncomouse discoveries and their internal development of Cre-lox technology, DuPont gained exclusive control over patents for these two technologies. In contrast to the University of Utah, DuPont exercised strict control over the distribution and use of mice that exploited the techniques covered by their patent portfolio. During the early 1990s, researchers (and their institutions) who wanted “freedom to operate” were obliged to obtain a license from DuPont when they sought to receive and use an Onco or Cre-lox mouse. The detailed licensing agreement required annual disclosures to DuPont regarding experimental progress, limits on informal mouse exchange among academic researchers, and “reach through” rights allowing DuPont to automatically receive licensing revenue from any commercial applications developed using either Cre-lox or Onco technology.

These limits to openness caused widespread discontent among the academic community. Academic researchers objected to the exercise of patent rights by a for-profit company as a significant limitation on the norms of openness among academics, and claimed that the lack of access to these mice significantly reduced their freedom to pursue their own research agendas.¹⁶Individual researchers engaged in various forms of protest – from attempt to initiate patent invalidation proceedings (which went nowhere) to informal sharing of mice (against the advice of their universities). As well, there were more systematic attempts to subvert or blunt the impact of the DuPont licensing regime: notably, in 1992 Dr. Ken Paigan, then director of the Jackson Laboratory (JAX), announced he would make Onco mice openly available without a license, directly contravening DuPont’s IP rights. While some researchers took advantage of informal sharing or access of Onco mice from the JAX (opening themselves to a potential infringement suit by DuPont), most researchers (and their institutions) were wary of the legal repercussions that could arise from using these mice, particularly for more applied research. Notably, through 1998, there was no access to Cre-lox mice through JAX or any other open-access depository.

Thus, by the late 1990s, researchers seeking to use a particular specialized

¹⁵Knock-out mice were covered under U.S. Patent 4,687737, Onco mice under U.S. Patent 4,736,866 and Cre-lox mice under U.S. Patent 4959317.

¹⁶As cited in Murray (2009), DuPont’s practices were seen as “an enormous obstacle to free and open distribution of information and materials. . . it was a whole new way of doing science. . . it really affected the way the mouse research community works”.

research mouse faced one of several access regimes. First, the most appropriate mouse for a particular research project might be a spontaneous mouse or a Knock-out mouse, and would (in general) be available on an open-access basis (from JAX or another provider) at marginal cost.¹⁷ Second, if the research required an Oncomouse, the mouse might be available informally through the peer-to-peer network or through JAX, but to use such a mouse (particularly for an applied project) was in direct contravention of DuPont’s licensing requirements. Third, if a Cre-lox mouse was preferred, it might be available through informal exchanges among colleagues. These informal exchanges were themselves beset by high transaction costs: Cre-lox developers invested considerable time and resources in its development and often required coauthorship (or other type of non-monetary payment) in exchange for access to their mice, and the exchange of such mice took place in the shadow of potential infringement suits (which meant contravening the official policy rules of most universities).¹⁸ It was also possible, in principle, to access Cre-lox and Onco mice by signing DuPont’s licensing agreement, though very few institutions or researchers signed an actual agreement prior to the NIH MoU. Finally, it was also feasible (at least in principle) to develop a new mouse as part of the research process, a process which could delay a project by at least 18 months and require significant resources and the development of specialized skills, and which could still be infringing on the DuPont patent portfolio.

3.3 The openness shocks on Cre-lox and Onco mice

The degree of openness associated with Cre-lox and Onco mice shifted dramatically in 1998 and 1999 respectively. Responding to considerable pressure from the academic community throughout the 1990s, the National Institutes of Health (NIH), with the direct involvement of NIH Director and Nobel Laureate Harold Varmus, successfully negotiated two Memorandum of Understanding (MoU) among DuPont, the Jackson Laboratories (JAX), and the NIH. Together, these two MoUs greatly increased the openness of genetically engineered mice for academic researchers. The Cre-lox MoU, announced in July 1998, allowed JAX or universities to distribute and share Cre-lox mice with a simple license (essentially a standardized one-page material transfer agreement and an institution-wide license). In addition, JAX announced its commitment to acquire, breed, and distribute Cre-lox mice on an open-access basis. A similar agreement for the Oncomouse was reached one year later (in July 1999), though the impact of this agreement was somewhat less dramatic as JAX had already been distributing Onco mice to researchers prior to the 1999 MoU.

Over a two-year period, life sciences researchers seeking to take advantage of

¹⁷In addition to the unenforced Utah patent on knock-out technology, a small number of additional patents were granted over specialized knock-out mice. However, the intellectual property restrictions associated with these mice seems to have been negligible, and, in any case, their openness was not directly influenced by the NIH agreements that we exploit in our empirical work.

¹⁸See Murray (2009).

the mouse genetics revolution thus experienced a significant shift in their ability to access and exploit Cre-lox and Onco mice, while experiencing no shift in the degree of openness for Knock-out and Spontaneous mice. These differences provide the key source of variation that we exploit in our empirical analysis. Three features are particularly useful to emphasize. First, while the “demand” for genetically engineered mice was increasing over time, there is no evidence that the potential demand for Onco or Cre-lox mice was increasing at a faster (or slower) rate than the demand for Knock-out mice. Each technology represented a general purpose research tool, with the key distinction being that the Knock-out technology was made available on an open-access basis throughout the period, while the Onco and Cre-lox technologies faced significant access restrictions until the time of the NIH agreements. Second, though the academic community lobbied continuously for increased openness regarding these research tools, there are good reasons to believe that the timing of the agreement as well as its scope were largely unanticipated.¹⁹ Thus it is unlikely that researchers delayed projects in anticipation of such a comprehensive agreement; instead, researchers deterred by licensing restrictions undertook different projects. Third, though the agreements cover two DuPont-controlled patents, they impacted a large number of specialized research mice. In spite of the IP difficulties, by 1998, more than 50 different engineered mice had been developed and disclosed in the scientific literature using the Cre-lox technology, and more than 160 different Onco mice were similarly described. As we outline in detail below, one can take advantage of the fact that these mice were developed and disclosed at different times and that their follow-on use by other scientists can be captured by the citation of these articles in follow-on scientific articles, in order to precisely identify the impact of the NIH openness agreements on the use of genetically engineered mice in follow-on scientific research.

4 Empirical strategy

4.1 Identification strategy

We examine the impact of shifts in the openness of some engineered research mice (arising from the NIH agreements) on the level and type of follow-on research. Building on Furman and Stern (2008), this approach addresses a fundamental inference problem associated with traditional cross-sectional approaches to the evaluation of openness (and related institutional arrangements) on scientific research: If more “open” inputs are used more extensively by follow-on researchers, does this follow from the fact that they are open or from the fact that openness tends to be associated with higher-quality inputs and materials? In the absence of an empirical framework that disentangles selection effects (i.e., the correlation between openness and overall research impact) from the marginal impact of openness *per se*, we cannot construct the appropriate counterfactual

¹⁹We discuss this point in more detail in Section 6.2 below.

estimate of the rate of follow-on research in the event that the same knowledge was available under a different level of openness.

Ideally, causal identification of the impact of openness would rely on a controlled experiment in which different knowledge inputs (such as particular research mice) were randomly allocated to distinct institutional environments with varying degrees of openness. A practical route capturing the essence of such an approach takes advantage of natural institutional variations that shift key research inputs towards higher (or lower) levels of openness in a way that is exogenous both to their initial production and to their incorporation into follow-on research lines.

We implement this idea by taking advantage of the institutional changes to openness negotiated by the NIH that affected some (but not all) research mice.²⁰ Our strategy exploits several distinctive elements of the system by which scientific research is disclosed and cited. First, new specialized research mice are disclosed through publication in scientific articles that describe their production and distinctive characteristics (we refer to these disclosures as mouse-articles). Notably, we are able to identify mouse-articles both for mice affected by the NIH MoU agreements (i.e., Cre-lox and Onco mouse-articles) and for mice unaffected by the NIH MoU agreements (i.e., Knock-Out and spontaneous mouse-articles).²¹ Second, we can trace out the impact of each mouse-article over time through the citations to that mouse-article by subsequent articles in the scientific literature. While an imperfect and noisy indicator of overall scientific impact, citations offer a systematic reflection of the process by which researchers acknowledge how their efforts at any one research stage build on the tools and knowledge developed by researchers in prior stages. More specifically, our approach focuses on the citation patterns associated with mouse-articles. Our qualitative research suggests that citations to a given mouse-article involve the use of that article's specialized research mouse in a follow-on experiment, and that most researchers routinely include a citation to the original mouse-article whenever a particular mouse is used in a follow-on project. Third, both the NIH agreements occurred well after the publication dates of the Cre-lox and Onco mouse-articles; thus for each mouse-article we are able to observe citations both before and after the NIH MoU (and compare this to the pattern observed for our control groups which were unaffected by the NIH MoUs). Finally, as noted above, both the timing and extent of the openness shock were largely exogenous. Specifically, the NIH agreement could have been reached, in principle, anytime from the early 1990s through the present. Moreover, our main control group – Knock-out mice – is likely to have been drawn from a population of

²⁰Our approach builds on recent work applying a differences-in-differences econometric framework to analyze the institutional and microeconomic foundations of knowledge accumulation (Murray and Stern, 2007; Furman and Stern, 2008; Huang and Murray, 2008; Rysman and Simcoe, 2008).

²¹While these types of mice differ in the precise details of the specialized genetic manipulation they allow, with the exception of Spontaneous mice, they are broadly similar in the scope of application and relevance to human disease. Moreover, all three were patented and could have been subject to strict enforcement. Spontaneous mice differ to the extent that they were not subject to patents.

similar scientific quality and importance, differing only insofar as the patent over Knock-out technology was unenforced by the University of Utah.²²

This empirical approach allows us to estimate pre- and post-MoU citation rates to the treated mouse-articles (those associated with Cre-lox and Onco mice). We also include untreated mouse-articles (Knock-out and Spontaneous mice) so as to more precisely identify a counterfactual estimate of the citation rate that would have occurred if the NIH agreement has not been signed. Overall, by measuring citations to Cre-lox and Onco mouse-articles before and after the openness MoUs, and by measuring the citations to mouse-paper articles unaffected by the MoUs, we can separately identify the causal impacts of the Cre-lox and Onco openness agreements.

4.2 Regression specifications

Our baseline regression takes the measure $Annual\ Citations_{jt}$ as its dependent variable, representing the number of citations to a given mouse-article j in a given calendar year t . On the RHS of the regression equation, we take $Post\ NIH\ MoU$ to be the key treatment variable, equal to one for mouse-articles impacted by an MoU for citation years after the MoU has had a chance to impact publication behavior. Finally, we take the variable $NIH\ MoU\ Window$ to be equal to one in the period between the signing of the agreement and the period when the MoU would have a chance to impact publication behavior.²³ Using a dataset composed of citations to mouse-articles impacted by the MoU and mouse-articles that are unaffected by the MoU, consider the following conditional fixed effects negative binomial estimator:

$$Annual\ Citations_{jt} = f(\varepsilon_{jt}; \gamma_j + \beta_t + \delta_{t-PubYear} + \Psi_0 NIH\ MoU\ Window_{jt} + \Psi_1 Post\ NIH\ MoU_{jt}), \quad (1)$$

where γ_j is an article fixed effect (conditioned out in estimation), β_t are citation-year effects and $\delta_{t-PubYear}$ are article-age fixed effects. These fixed effects account for the heterogeneity among scientific articles, the nonlinear evolution of citations over time elapsed since initial publication, and the potential for differences over time in citation practices. This specification also accounts for the incidental parameters problem (Hausman, Hall and Griliches, 1984), testing for the impact of the NIH agreements by estimating how the citation rate for a mouse-article *changes* in response to the NIH MoU, accounting for fixed differences in the citation rate across articles and relative to the non-parametric trend in citation rates for the non-treated control groups.

We then turn to evaluating the impact of the openness shocks on different types of citations. We use the citations from each citation year to tabulate

²²We find support for this view in our analysis of pre-MoU trends in Section 6.2 and Table 7, where we show that prior to their respective MoU dates, Cre-lox and Onco mice have citation flows that are statistically indistinguishable from those of Knock-out mice.

²³Consistent with our discussion in Section 3, the window period for Cre-lox mice covers 1998 and 1999, and the window period for Onco mice covers 1999 and 2000.

key characteristics into two mutually exclusive types, and estimate the impact of the NIH MoU on each citation-year margin. For example, the model predicts that openness should increase the number of distinct researchers utilizing a given specialized research mouse. To test this hypothesis we estimate the difference of the impact of a shift in openness on the number of authors publishing follow-on research who have previously cited a particular mouse-article (*Old Authors_{jt}*) versus those who have not previously cited a particular mouse-article (*New Authors_{jt}*).

More specifically, we jointly estimate the two equations:

$$\begin{aligned} \text{New Authors}_{jt} = & f(\varepsilon_{jt}; \gamma_j + \alpha t + \beta_t + \delta_{t-\text{PubYear}}^{\text{NEW}} \\ & + \Psi_0^{\text{NEW}} \text{NIH MoU Window}_{jt} \\ & + \Psi_1^{\text{NEW}} \text{Post NIH MoU}_{jt}) \end{aligned} \quad (2)$$

and

$$\begin{aligned} \text{Old Authors}_{jt} = & f(\varepsilon_{jt}; \gamma_j + \beta_t + \delta_{t-\text{PubYear}}^{\text{OLD}} \\ & + \Psi_0^{\text{OLD}} \text{NIH MoU Window}_{jt} \\ & + \Psi_1^{\text{OLD}} \text{Post NIH MoU}_{jt}), \end{aligned} \quad (3)$$

where γ_j is a mouse-article fixed effect, α parametrizes a linear calendar-time-trend difference between the two equations, β_t is a calendar-time fixed effect, and $\delta_{t-\text{PubYear}}^{\text{NEW}}$ and $\delta_{t-\text{PubYear}}^{\text{OLD}}$ are article-age fixed effects. To evaluate whether the change in citations occurring after the openness shock is concentrated in citations by authors who had not previously cited a particular mouse-article, we test whether $\Psi_1^{\text{NEW}} > \Psi_1^{\text{OLD}}$. This specification includes several parametric restrictions, including setting the mouse-article fixed effects γ_j and calendar-time fixed effects β_t to be equal across the two equations (2) and (3), and imposing a linear functional form (parametrized by α) on the difference in the effect of calendar time across the two equations. We do allow for the publication-age fixed effects to vary freely across (2) and (3), as the evolution of citations in the time elapsed since publication will differ significantly for the two citation margins (in particular, most citations in the first few years after publication will be associated with “new” authors).

We then undertake a similar exercise to evaluate whether a boost in citations is associated with (a) new versus old institutions, (b) new versus old key words, and (c) new versus old journals. Finally, we explore the research response to the openness shocks along a given research line by comparing citations in applied versus basic journals.

This empirical framework also allows us to test whether citations to mouse-articles in both the treatment and control groups evolve in a similar way over time, except for shifts in the institutional environment. We can test this assumption directly by allowing for a time trend specific to the treatment group

for each citation margin. Because different mouse-articles are published in different years we are able to disentangle the treatment effect of the NIH MoUs from secular differences in the citation trends of articles in the treatment group. At the same time, our theoretical discussion suggests that the treatment effect should grow over time i.e. with the time elapsed from the openness shock (MoU). We thus include a specification which separately estimates the short-term and long-term impacts of the MoU. Thus when testing for a treatment group specific trend we will separately allow for the treatment effect itself to change in the time following the treatment. Lastly, it is also possible to test whether there is an unanticipated increase in citations in the periods immediately preceding the MoU, thereby suggesting that the timing of the MoU was endogenous. We explore this possibility by testing for a pre-MoU period in the years just prior to the signing of the MoU.

5 Data

5.1 Data and sampling

The data for this study are drawn from the entire population of research mice catalogued by the Mouse Genome Informatics (MGI) database. MGI consists of over 13,000 unique mice, each of which can be linked to a publication in the scientific literature describing its initial disclosure, thereby establishing a population of mouse-articles. Of this large population, we focus only on mouse-articles published between 1987 and 1998 (the date of the first NIH agreement). As outlined above we sample all mouse-articles for the four types of mouse engineering technologies defined by MGI: Cre-lox (28), Onco (102), Knock-out (1895) and Spontaneous (146). Our sample thus includes 2171 novel mice, each linked to a unique mouse-article.

We use PubMed and Thomson ISI Web of Science to collect detailed bibliometric information on all follow-on forward citations in academic journals through 2006. Each of these 432,083 citations includes information on last author, reprint author, institutional addresses, key words, and journal characteristics (including journal name, journal impact factor and a score for basicness). Citations are then aggregated into 22,265 citation-year observations by combining all the citations received by a given mouse-article in any particular year as the basis for our analysis.

To capture the *composition* of follow-on research, we code citation characteristics into a set of mutually exclusive categorical variables. We focus on margins intended to capture the degree of horizontal experimentation across new lines, as well as vertical exploitation. To illustrate the construction of these variables, take the case of new key words. For each forward citation, ISI Web of Science provides a series of keywords (referred to as Key Words Plus). We first take the list of all key words in the set of citations a particular mouse-article receives in a given year, and remove all duplicates to obtain a list of unique key words associated with that citation-year observation. We then categorize a given key word

to be new if it has never been used in citations to a particular mouse-article in any prior year. We code as old all key words that have appeared in citations to the particular mouse-article in prior years. This construction allows us to capture changes in the research landscape over time. We generate four new/old categorical variables:

- i. New/Old Last Author: defined as new if the last author has never appeared as a last author before in a citation to the mouse-article in prior years, old otherwise.
- ii. New/Old Institution: defined as new if an address in the institution list has never appeared in an address list of citations to the mouse-article in prior years, old otherwise.
- iii. New/Old Key Words: defined as new if a key word has never before appeared in the key word list of citations to the mouse-article, old otherwise.
- iv. New/Old Journal: defined as new if the journal of the citation has never appeared before in the citations to the mouse-article, old otherwise.

We also categorize citations according to whether they are published in basic or applied journals.²⁴ This allows us to capture the predictions of our model regarding the impact of openness on the vertical direction of follow-on innovation, i.e. whether these shifts lead to research further along particular research lines (towards commercialization). It is worth noting that in this analysis, multidisciplinary journals are classified as “basic”.

These measures reflect various ways in which openness impacts on follow-on research along both horizontal and vertical dimensions. Using the two-equation framework described in Section 4, they allow us to test the hypothesis that changes in openness create more diverse lines of research, pursued by a more diverse range of scientists. We also investigate whether openness is associated with more basic or applied research.²⁵

5.2 Variables and summary statistics

Table 1 provides variable names and definitions and Table 2 reports summary statistics. The dependent variable in the initial set of regressions is *Annual Citations_{jt}* which measures the total number of citations received by mouse-article j in year t . The average of *Annual Citations_{jt}* is 19.41 (with a minimum of 0 and maximum of 336), highlighting the overall importance of mouse genetics research in this period. Because we observe citation-years

²⁴Our Basic/Applied Journal definition is based on work by Lim (2004) who used the measure building on a classification scheme developed by CHI Research, Inc. According to Lim, “CHI awards each journal a score from zero to four. For the biomedical sciences, they correspond to clinical observation, clinical mix, clinical investigation and basic science.”

²⁵It is worth noting that we do not examine the impact of openness on the academic/industry citation margin. The NIH MoUs were directed specifically to public-sector researchers and 97.5% of all forward citations have at least one of their authors in public institutions.

from 1993 through 2006, the average $Citation\ Year_{jt}$ is 2001. We also create an alternative dependent variable $High\ Quality\ Ctations_{jt}$, which captures the production of research which appears in a “top 50” journal (with mean equal to 4.3). We then construct a series of dependent variables based on the key categorical margins of interest:²⁶

- i. $New\ Last\ Authos_{jt}$ and $Old\ Last\ Authors_{jt}$, with mean values equal to 11.7 and 3.9 respectively.
- ii. $New\ Institutions_{jt}$ and $Old\ Institutions_{jt}$, with mean values equal to 17.5 and 10.2 respectively.
- iii. $New\ Keywords_{jt}$ and $Old\ Keywords_{jt}$, with mean values equal to 74.9 and 55.4 respectively.
- iv. $New\ Journals_{jt}$ and $Old\ Journals_{jt}$, with mean values equal to 7.9 and 6.2 respectively.
- v. $Basic\ Citations_{jt}$ and $Applied\ Citations_{jt}$, with mean values equal to 9.2 and 7.4 respectively.

TABLE 1 HERE

TABLE 2 HERE

We then define two measures that will be used to estimate the impact of the NIH MoUs. We divide the period after the NIH MoU signing into two subperiods because the NIH agreements likely take time to influence follow-on research. Specifically, we define a window period and a treatment period to allow for a reasonable lag (2 years) for the NIH MoU to impact observed publication patterns. $NIH\ MoU\ Window$ (mean equal to 0.011) is a dummy variable equal to one for articles impacted by that MoU during the year where the NIH MoU was signed and the following year (1998/1999 for Cre-lox mouse-articles, 1999/2000 for Onco mouse-articles). $Post\ NIH\ MoU_{jt}$, (mean equal to 0.036), our key treatment variable, is a dummy variable equal to one for all articles impacted by the MoUs in citation-years after the window period ended. We also define specific treatment variables for each of the MoU agreements: $Cre\ Lox\ MoU\ Window_{jt}$, $Post\ Cre\ Lox\ MoU_{jt}$, $Onco\ MoU\ Window_{jt}$ and $Post\ Onco\ MoU_{jt}$.

To examine the short-term versus long-term impact of the NIH MoU, we also define a treatment variable, $Post\ NIH\ MoU, Short\ Term$ equal to one

²⁶Note that the sum of the annual means of each margin need not add up to the mean annual citation count. First, due to data-matching issues we cannot always identify 100% of citations as belonging to one or the other margin; this leads to a sum lower than the mean annual citation count. Second, new/old margins focus on the count of unique instances of the characteristic in question; for example, if there are multiple citations from a particular journal to a mouse-article in a given year, we only count the first such citation. This also leads to a sum lower than the mean annual citation count. Finally, for the counts of institutions and keywords, each citation contains multiple entries for these fields, leading to counts higher than the mean annual citation count. For example, in the case of keywords, the sum of the margin means is just over 120, indicating that the average paper has between 6 and 7 keywords.

for the first three years after the window period, and a separate measure, *Post NIH MoU, Long-Term*, for mouse-articles impacted by the NIH MoUs for the period four years or more after the window.

We highlight our summary statistics disaggregated by the type of mouse technology in Table 3. The most salient point to note is that compared to the overall sample mean of 18, the *Annual Citations_{jt}* for Cre-lox and Onco mice are 15 and 12 respectively. However, the Spontaneous mice in our control group have a lower mean (4), the Knock-out mice have mean *Annual Citations_{jt}* of over 21, thus providing further substantiation of the comparability of the treatment and control groups. Moreover, the mean publication year and mean number of authors across the four mouse technologies are similar.

TABLE 3 HERE

6 Results

Now we proceed to estimate the causal impact of the MoU openness shocks on the overall flow of citations (Table 4), and then turn to the core of our analysis which examines the impact of the NIH MoUs on the composition of citations. Specifically, we examine the impact of openness on the type -new versus old- of researchers (Table 5) and the nature -new versus old- of research directions (Table 6). Moreover, we undertake several robustness checks including an analysis that specifically allows for differences in the time trend of citations for our treatment and control groups (Table 7). Lastly, we analyze the impact of openness on the vertical exploitation of particular research lines by examining the composition of follow-on research in basic versus applied journals (Table 8). By adopting a differences-in-differences approach throughout our analysis to evaluate the impact of openness on different citation margins, we are able to infer the relationship between openness and academic freedom.

In all our Tables we report coefficients estimates and their corresponding incidence-rate ratios (IRRs). We discuss our results in terms of IRRs because they are easily interpreted as percentage changes relative to a baseline (i.e. the null hypothesis of no effect yields a coefficient of 1.0). As well, all of the models report block bootstrapped standard errors clustered by mouse-article (MacKinnon, 2002).

6.1 Impact of openness on the level of follow-on research

Our regression results begin in Table 4 with a conditional fixed effect negative binomial specification using *Annual Citations_{jt}* as the dependent variable. All specifications also include the full set of article, age and calendar year fixed effects. In (4-1) we include both the *NIH MoU Window* and the *Post NIH MoU* regressors. The results are striking: After accounting for the window period, mouse articles impacted by an MoU experience a 30% increase in their annual citation rate. As illustrated in (4-2), the impact of

the NIH MoUs are increasing over time: While the increase in citations in the three years after the window period is equal to 22%, the coefficient on *Post NIH MoU, Long-Term* suggests that the permanent effect is nearly doubled (at 43%). Not simply a reflection of publication lags, the results in (4-2) suggest the presence of a positive and permanent increase in the use of genetically engineered mice which have been shifted to a higher level of openness. In (4-3) we estimate separate coefficients for the Cre-lox and Onco MoUs: Both are statistically significant although the magnitude of the boost to citations associated with the Cre-lox MoU is larger (47% for citations to Cre-lox mouse-articles compared to 27% for citations to Onco mouse-articles). Finally, in (4-4), we undertake a robustness check by focusing on citations in “high impact” journals. We find strong evidence for a 41% boost in high quality citations suggesting that the impact of the shift in openness is concentrated in research appearing in the most prestigious journals.

The results in Table 4 provide strong support for a central hypothesis – that positive shocks to openness foster follow-on research. These findings reinforce previous studies of the impact of openness and accessibility such as Furman and Stern (2009) and Murray and Stern (2007). Furthermore, our results are consistent with a multi-staged view of innovation: an increase in openness does not simply lead to a temporary increase in follow-on research but instead has an increasing impact over time. Finally, though we hold off on this discussion until Table 7, we can show that the estimated impact of the NIH MoUs are not simply due to a different time trend for the treatment and control groups. Taken together, these results highlight the sensitivity of follow-on researchers to the degree of openness of critical research inputs.

TABLE 4 HERE

6.2 Impact of openness on the type of follow-on research: horizontal exploration

Tables 5 and 6 present our main evidence regarding the theoretical claim that greater openness results in greater horizontal experimentation, spawning a more diverse array of research lines and encouraging the participation of new researchers. In Table 5, our key comparison is between researchers listed as the last author (senior scientist) on citations who have (or have not) been previously listed on a citation to the mouse-article of interest, as captured in our measures *New Authors* and *Old Authors*. In (5-1a) and (5-1b) we estimate whether the marginal impact of *Post NIH MoU* is different for new versus old last-authors. The results are consistent with the predictions from the control approach: while there is only an insignificant 13% increase in citations by old authors, the increase by new authors is estimated to be significant and more than 38%. Moreover, these two coefficients are significantly different from each other. We then estimate a separate coefficient for the short-term versus long-term impact of the NIH MoUs on new versus old authors (5-2a and 5-2b). The increase in citations by new authors is greater than the increase in citations for

old authors in both the short- and long-term (with the difference between the two coefficients being significant at the 1% level). Strikingly, the estimate of the long-term increase in new author citations is greater than 50%. Moreover, when we separately estimate the impact of the Cre-lox and Onco MoUs on new versus old authors, (5-3a and 5-3b) we find that the estimated boost for new authors is statistically significant for each agreement compared to a much smaller and statistically insignificant increase in citations by old authors. Moreover, we find that the difference between the new versus old coefficients is significant for each agreement at the 5% level.

Finally, in (5-4a) and (5-4b) we turn to an alternative measure of the diversity of follow-on researchers as captured by their institutional affiliation. If researchers within a given institution (e.g. Northwestern University) share mice with each other, an increase in openness would be associated with an increasing number of institutions making investments in particular specialized research mice. Conversely, any university-level agreement made prior to the MoU would allow for follow-on research by all scientists within that specific institution. Similar to the results for new versus old authors, the boost in citations associated with the NIH MoUs is concentrated in citations from institutions that had not previously cited that mouse article (27% vs. 13% boost).

Overall, the results in Table 5 provide direct evidence that the shift in openness associated with the NIH MoUs expanded the diversity of researchers drawing on a particular line of research. In other words, these findings are consistent with the idea that an increase in openness reducing the fixed cost of critical upstream inputs expands the range of researchers willing to undertake such investments in exploratory research areas.

TABLE 5 HERE

We then turn in Table 6 to the related prediction that openness enhances the diversity of research lines (particularly in an academic research environment where scientists are free to choose their own research direction). We capture the degree of horizontal diversity by using the key words that categorize each citation (recall that key words are chosen by the archiving service rather than the researchers). In (6-1a) and (6-1b) we compare the impact of the NIH MoUs on *New Key Words* and *Old Key Words* respectively. While there is a small and statistically insignificant decline in the number of old key words there is a significant 26% increase in the number of citations with new key words. Moreover, these coefficients are statistically significantly different from each other. This is not just a short-term effect: the analysis of time dynamics in (6-2a) and (6-2b) indicates that there is an even larger 41% increase in the number of new key works in the long-term, relative to an insignificant decrease to old key words in the long-term; moreover, the difference between this 41% increase and the 18% increase to new key words in the short-term is statistically significant at the 1% level. When we decompose the openness changes into the Crelox and Onco MoUs (see 6-3a and 6-3b), we continue to find a quantitatively and statistically significant difference between the new and old key words

coefficients. Both the Cre-lox and Onco MoU are associated with a significant boost in new key words (40% and 21% respectively) and a small and insignificant decline in old key words. Finally, as in our analysis of the diversity of citing researchers, we use an alternative measure to test the robustness of our findings on research diversity. In (6-4a) and (6-4b) we compare the citation margins between *New Journals* and *Old Journals*, where a "new" journal is one which has never before published an article citing the original mouse-paper article in question. We find that being in the *Post NIH MoU* period leads to a 38% increase (significant at the 1% level) in citations from new journals, and only a 20% increase in citations from old journals (significant at the 5% level).

TABLE 6 HERE

In our analysis so far we have assumed that the citation-age profile is similar for the treatment and control groups. Because different mouse articles are published at different times (relative to the MoUs) we can test this assumption directly. In Table 7 we re-estimate each of the key equations for overall citations, new versus old authors, and new versus old key words, allowing for a time trend specific to the treatment group for each citation margin. Of course, since the treatment effect itself is predicted to increase in the time elapsed since the treatment we separately allow for a post-MoU trend. The results reinforce our overall findings. First and most importantly, across all of the specifications there is no statistically significant or quantitatively important trend specific to the treatment articles. Second, the estimated coefficients for the impact of the MoU are consistent with our earlier findings, although the coefficients are smaller (since they are implicitly estimating the impact of the MoU only for the first year after the window period). More importantly, there is a significant impact of the treatment over time for overall citations, new authors and new key words. While there is also an increase over time for old authors and old key words, the coefficient is smaller and noisier.

We have also experimented extensively with specifications that estimate coefficients on a year-by-year basis relative to the time of the MoU, in order to test for the presence of a pre-shock trend in either of the treatment groups (relative to controls) and to examine the evolution of each citation margin after the shock. While the pre-deposit trend is not statistically significant for any of the citation margins we consider, it is also true that these year-by-year coefficients are imprecisely estimated, in part because of the relatively small number of annual citation-year observations in the treatment groups. In other words, there is no evidence for a significant increase in citations prior to the MoU that might raise concerns about the endogeneity of the timing of the agreement.

TABLE 7 HERE

6.3 Impact of openness on the type of follow-on research: vertical exploitation

In Table 8 we turn to the effects of openness shocks on the vertical distribution of research. We do so by examining the marginal impact of the openness shocks on the production of research in basic versus applied research journals. In (8-1a) and (8-1b), we find that the *Basic Citations* dependent variable increases by 26% during the post-MoU period; at the same time, *Applied Citations* experience a 30% increase during that period (both significant at the 1% level). This suggests that the overall impact of the MoUs involves both basic and applied citations. We then disentangle the separate impacts of the Cre-lox and Onco MoUs. Recall that in the pre-MoU period, not only was there stringent reach-through rights associated with Cre-lox mice, but also very limited access to the mice due to ex ante enforcement by DuPont. In contrast, Onco mice were made available in the pre-MoU period through the Jackson Labs. As a result, relative to the Cre-lox MoU the Onco MoU reduced reach-through rights but had a more limited impact on access. In (8-2a) and (8-2b) we evaluate the differential impact of these MoUs on basic versus applied citations. We find that the impact of the Cre-lox MoU is concentrated in basic citations, while the Onco shock has a significant effect only on applied citations. Specifically, the Cre-lox MoU leads to a 120% increase in basic citations (significant at the 1% level) but no change in applied citations, the Onco MoU leads to a 57% increase in applied citations and has no significant impact on basic citations. This is consistent with the view that when upstream access is already secured (as was the case for Onco mice), then an agreement that shifts the balance of appropriability toward follow-on innovators induces more applied research.

TABLE 8 HERE

6.4 Impact of openness on the creation of new genetically modified mice

Finally, we analyze the impact of the Cre-lox and Onco mice openness shocks on the rate of creation of genetically modified mice. Our empirical approach begins with the full set of mice in the MGI database (approximately 13,000). For each mouse and its associated mouse-article, we obtain information regarding the technology used in its creation, and the journal in which the article was published. We focus on mice created since 1983, using one of the four technologies present in our citation analysis. Figure 1 describes the creation of mice by technology category, from 1990 through 2006, in order to provide a visual representation of any potential impact on mouse creation as a result of the MoUs negotiated by the NIH for Cre-lox and Onco mice. A standard appropriability story would predict a significant reduction in the creation of Cre-lox and Onco mice, relative to the control technologies, following their corresponding NIH-MoU agreements. However, rather than a sharp decrease, the graph does not exhibit any marked change in creation rates around the time of the two

MoUs (1998 and 1999, respectively). Thus, a simple visual analysis of mouse creation does not offer support for the simplest version of the appropriability perspective.

FIGURE 1 HERE

Our next step is to test our model’s predictions regarding mouse creation using a differences-in-differences regression setup similar to the one in our citation analysis. Specifically, our baseline regression takes the measure *Annual Mouse Creation_{kt}* as its dependent variable, representing the number of mouse-articles published to a given technology *k* in a given calendar year *t*. Our key regressors are interaction effects between each of the MOU agreements and the Annual Mouse Creation years impacted by those agreements: *Cre-Lox MoU Window_{kt}*, *Post Cre-Lox MoU_{kt}*, *Onco MoU Window_{kt}* and *Post Onco MoU_{kt}*. We also include controls for calendar-year, age, and technology-specific effects. Similar to our earlier specifications, we include a technology fixed effect (conditioned out in the context of our conditional fixed effect negative binomial estimator), calendar year fixed effects, and a cubic polynomial in the age of the technology. Finally, because of the much older age of the Spontaneous mice, we include a separate linear time trend for the Spontaneous group. This results in the following specification:

$$\begin{aligned}
 \text{Annual Mouse Creation}_{kt} = & f(\varepsilon_{kt}; \gamma_k + \beta_t + g(t - \text{EarliestYear}_k) & (4) \\
 & + \sigma I(k = \text{“Spontaneous”}) * (t - \text{EarliestYear}_k) \\
 & + \Psi_{0,1} \text{Cre-Lox MoU Window}_{kt} \\
 & + \Psi_{1,1} \text{Post Cre-Lox MoU}_{kt} \\
 & + \Psi_{0,2} \text{Onco MoU Window}_{kt} \\
 & + \Psi_{1,2} \text{Post Onco MoU}_{kt}),
 \end{aligned}$$

Overall, this specification tests for the impact of the NIH agreements by estimating how the mouse-creation rate for a technology *changes* in response to the NIH MoU, accounting for fixed differences in the creation rate across technologies and relative to the trend in creation rates for the non-treated control groups.

Our results of this differences-in-differences specification are in the first two columns of Table 9. In (9-1), we use the four technologies we studied in our citation analysis; the estimates suggest that, rather than a decrease (per the appropriability story), there was a 50% increase in mouse creation after the MoUs for both Cre-lox and Onco mice. However, this estimated increase is a fragile result. In (9-2), we expand the control group to include all major genetic modification technologies in the MGI database, adding the following four technologies: Targeted(reporter), Targeted(knock-in), Targeted(Floxed/Frt), and Genetrapped mice. The strong increase in mouse creation reported in (9-1) largely disappears when we include a wide set of control groups. For Onco, we

now estimate virtually no change in mouse creation; for Cre-lox, we still find an increase of over 20% in mouse creation, but this result is no longer statistically significant. However, both (9-1) and (9-2) offer an important take-away: *increased openness does not lead to a reduction in overall new mouse creation*. This contradicts the main prediction of the appropriability view, where one consequence of the increased openness on upstream research implied by the MoUs would be to discourage the development of new mice. Instead, consistent with the control view, a more open environment had either a neutral or positive impact on the creation of new mouse varieties.

Finally, we also consider the impact of the openness shocks on the diversity of mice being created. Under the control approach, openness is likely to increase the number of distinct journals (and, correspondingly, research areas) in which mice from a given GM technology are published (whereas the appropriability view predicts a decline in all types of new mice creation). To evaluate these competing hypotheses, we divide the mice created through a particular technology in a given year from each citation year into two mutually exclusive types: those published in journals which had previously published mice associated with that technology (“Old Creation Journals”), and those which have not previously published any such mice (“New Creation Journals”). Our specifications here are simply a two-equation version of the specifications employed in (9-1); we estimate the impact of each of the MoUs on new mice published in old versus new journals, accounting for fixed differences across each technology type, calendar year effects, and including time trends as above. The results are in (9-3) and (9-4). We find economically and statistically significant (positive) differences between the number of “new journals” and “old journals” in which new mouse-articles are published for the window periods of both Cre-lox and Onco mice, and during the treatment period for Onco mice.

Overall, the results in Table 9 show that there is no reduction in the creation of genetically modified mice following the NIH-MoU agreements, but that there is a strong increase in the diversity of mouse creation, as measured by the journals in which mouse-articles are published. We interpret these results to cast some doubt on the ability of the appropriability story to fully explain the process of innovation in our setting, and to lend credence to our control story and its emphasis on horizontal exploration in academic research.

TABLE 9 HERE

7 Conclusion

In this paper we argued that increased openness of upstream research raises the incentives for additional upstream research by encouraging the establishment of entirely new research directions. We tested this hypothesis by examining a “natural experiment” in openness within the academic community: NIH agreements during the late 1990s that circumscribed IP restrictions for academics and increased the openness of key types of genetically engineered mice and the research tools associated with their production.

Our empirical results suggest that the NIH MoUs had a profound and long-lasting impact on follow-on research. Moreover, as predicted by the control approach, we find that these openness shocks not only increased the overall flow of follow-on research using specific engineered research mice, but that they also expanded the diversity of researchers working on particular research lines, and the diversity of the research taking place. Finally, we find that increased openness does not reduce the flow of new mouse creation. Overall, our results speak to a world in which increased openness, stemming from a reduction in IP restrictions, increases the availability and access to multiple research tools, where these tools can in turn be used in different ways in subsequent research, which can itself generate new lines by cross-fertilization. Our results therefore highlight a key limitation in the current literature on intellectual property and innovation - the potential restrictions intellectual property rights may place on the diversity of research and researchers who would otherwise take a single powerful idea and experiment across multiple research lines.

The analysis developed in this paper could be extended in several interesting directions. One avenue would be to reassess the Bayh-Dole Act based on our findings. Indeed, our results highlight one of the possible dangers of excessive IP enforcement: namely, if IP is used to restrict openness at very early stages of the research line, then it may stifle exploratory projects that are necessary for diverse follow-on innovation. Further, our analysis suggests that more attention be paid by economists to recent corporate attempts to generate new sources of profit through building on the openness of knowledge production by others. Tapscott and Williams (2006) explain how IBM has recovered from competition with Microsoft by engaging in the openness promoted by the Linux community. By contrast, the experience of DuPont and other companies that kept on enforcing patents while also attempting to engage with the open scientific community²⁷ was less successful. This pattern suggests the need for a systematic analysis of the forces and trade-offs at work in an economic environment where both proprietary-technology and open-technology firms compete with each other and cooperate with open communities, setting the stage for future research.

²⁷See Huang and Murray (2008).

References

- [1] Aghion, P, Dewatripont, M, and Stein, J (2008), “Academic Freedom and Private Sector Focus” , mimeo, Harvard, forthcoming in *The Rand Journal of Economics*.
- [2] Arrow, K (1962), “Economics of Welfare and the Allocation of Resources for Invention”, in R. Nelson ed., *The Rate and Direction of Inventive Activity*, Princeton University Press.
- [3] Azoulay, P, Waverly, D, and Stuart, T (2007), “The Impact of Academic Patenting on (Public) Research Output”, *Journal of Economic Behavior and Organization*, 63 (4), 599-623.
- [4] Bresnahan, T, and Trajtenberg, M (1995), “General Purpose Technologies: ‘Engines of Growth’?,” *Journal of Econometrics*, 65(1), 83-108
- [5] Dasgupta, P and David, P (1994), “Toward a New Economics of Science” , *Research Policy*, 23, 487-521.
- [6] David, P (2001a), “From Keeping ‘Nature’s Secrets’ to the Institutionalization of ‘Open Science’”, Stanford Working Paper #01-006.
- [7] David, P (2001b), “Will Building Good Fences Really Make Good Neighbors in Science”, Stanford Working Paper #01-005.
- [8] David, P (2003), “The Economic Logic of ‘Open Science’ and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer”, SIEPR Discussion Paper #02-30.
- [9] Furman, J, and Stern, S (2006), “Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research”, Working paper, National Bureau of Economic Research, Cambridge, MA.
- [10] Gallini, N, and Scotchmer, S (2002), “Intellectual Property: When Is It the Best Incentive System?”, *Innovation Policy and the Economy*, 2, 51-77.
- [11] Green, J, and Scotchmer, S (1995), “On the Division of Profit in Sequential Innovation”, *The RAND Journal of Economics*, 26, 20-33.
- [12] Hall, B, Jaffe, A, and Trajtenberg, M (2001), “The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools”, Working paper no. 8498, NBER, Cambridge, MA.
- [13] Hausman, J, Hall, B, and Griliches, Z (1984), ”Econometric Models for Count Data with an Application to the Patents-R & D Relationship”, *Econometrica*, 52, 909-938 .
- [14] Heller, M (2008), *Gridlock Economy: How Too Much Ownership Wrecks Markets, Stops Innovation and Costs Lives*, Philadelphia PA: Basic Books.

- [15] Heller, M, and Eisenberg, R (1998), "Can Patents Deter Innovation? The Anticommons in Biomedical Research", *Science*, 280, 698-701.
- [16] Henderson, R, Jaffe, A, and Trajtenberg, M (1998), "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988", *Review of Economics and Statistics*, 80, 119-127.
- [17] Howitt, P (2003), "The Economics of Science and the Future of Universities", The Timlin Lecture 2000, University of Saskatchewan.
- [18] Huang, K, and Murray, F (2008), "Does Patent Strategy Shape The Long-Run Supply Of Public Knowledge? Evidence From The Human Genome", forthcoming in *Academy of Management Journal*.
- [19] Kremer, M (1998), "Patent Buyouts: A Mechanism for Encouraging Innovation", *The Quarterly Journal of Economics*, 113, 1137-1167.
- [20] Kortum, S, and Lerner, J (2000), "Assessing the Contribution of Venture Capital to Innovation", *The RAND Journal of Economics*, 31, 674-692.
- [21] Lim, K (2004), "The Relationship Between Research and Innovation in the Semi-Conductor and Pharmaceutical Industries", *Research Policy*, 33, 287-321.
- [22] Merges, R and Nelson, R (1990), "On the Complex Economics of Patent Scope", *Columbia Law Review*, 90, 839-916.
- [23] Merton, R (1973), *The Sociology of Science: Theoretical and Empirical Investigations*, Edited by Norman Storer. Chicago: University of Chicago Press.
- [24] Mokyr, J (2002), *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton: Princeton University Press.
- [25] Mowery, D, Nelson, R, Sampat, B and Ziedonis, A (2001), "The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980", *Research Policy*, 30:99-119.
- [26] Murray, F, and 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 and Organization*, 63:4, 648-687.
- [27] Murray, F (2002), "Innovation as Co-Evolution of Scientific and Technological Networks: Exploring Tissue Engineering", *Research Policy*, 31, Issues 8-9, 1389-1403.
- [28] Murray, F (2007), "The Stem Cell Market: Patents & the Pursuit of Scientific Progress", *New England Journal of Medicine*, 356 (23), 2341-2343.

- [29] Murray, F (2009), “The Oncomouse that Roared: Hybrid Exchange Strategies as a Source of Productive Tension at the Boundary of Overlapping Institutions”, forthcoming in the *American Journal of Sociology*, and Working Paper, Sloan School of Management.
- [30] Nelson, R (1959), “The Simple Economics of Basic Scientific Research,” *Journal of Political Economy* 67(3), 297-306.
- [31] O’Mahony, S (2003), “Guarding the Commons: How Community Managed Projects Protect Their Work”, *Research Policy*, 32, 1179-1198.
- [32] Rader, K (2004), *Making Mice*, Princeton, NJ: Princeton University Press.
- [33] Rosenberg, N and Trajtenberg, M (2004), “A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States”, *The Journal of Economic History*, 64, 61- 99.
- [34] Rysman, M, and Simcoe, T (2008), “Patents and the Performance of Voluntary Standard Setting Organizations”, forthcoming in *Management Science*, and Working Paper No. 05-22, NET Institute. Available at SSRN: <http://ssrn.com/abstract=851245>.
- [35] Scotchmer, S (1991), “Standing on the Shoulders of Giants: Cumulative Research and the Patent Law”, *Journal of Economic Perspectives*, 5, 29-41.
- [36] Scotchmer, S (1996), “Protecting Early Innovators: Should Second-Generation Products be Patentable?”, *The Rand Journal of Economics*, 27, Summer 1996, 322-331.
- [37] Scotchmer, S (2004), *Innovation and Incentives*, Cambridge, MA: MIT Press.
- [38] Stern, S (2004), ”Do Scientists Pay to Be Scientists”, *Management Science*, 50, 835-853.
- [39] Tapscott, D, and Williams, A (2006), *Wikinomics: How Mass Collaboration Changes Everything*, New York, NY: Portfolio Press.
- [40] Williams, H (2009), “Intellectual Property Rights and Innovation: Evidence from the Human Genome”, Working Paper, Department of Economics, Harvard University.

FIGURE 1: MOUSE CREATION BY TECHNOLOGY

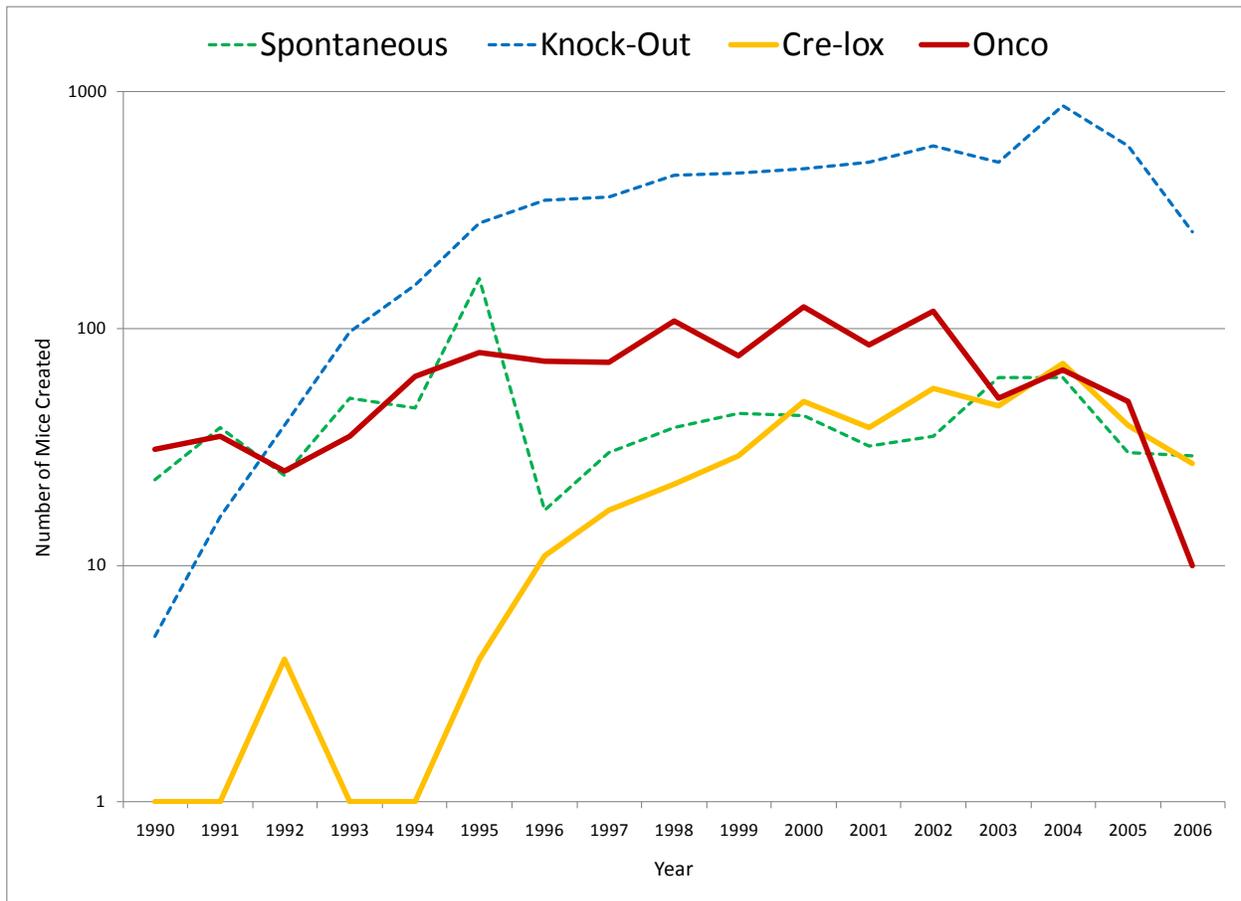


TABLE 1: VARIABLES & DEFINITIONS

VARIABLE	DEFINITION	SOURCE
PUBLICATION CHARACTERISTICS		
Publication Year	Year in which Article j is published	PM
# Authors	Count of the number of authors of Article j	PM
Total Citations	# of FORWARD CITATIONS from publication date through 2006	SCI
CITATION-YEAR CHARACTERISTICS		
Annual Citations	Count of all Forward Citations to Article j in Year t	SCI
High Quality Annual Citations	Count of Forward Citations to Article j in Year t where the journal of the citation is a top 50 journal based on impact factor rankings.	ISI
Citation Year	Year in which FORWARD CITATIONS are received	SCI
CITATION-YEAR MARGIN CHARACTERISTICS		
Basic Citations	Count of Forward Citations to Article j in Year t where the journal of the citation is a basic-research journal	CHIBasic
Applied Citations	Count of Forward Citations to Article j in Year t where the journal of the citation is an applied-research journal	CHIBasic
New X	Count of unique values of characteristic X of Forward Citations to Article j in Year t which are “new” and have not appeared in the citations to Article j in prior years.	
Old X	Count of unique values of characteristic X of Forward Citations to Article j in Year t which are NOT “new” and <i>have</i> appeared in the citations to Article j in prior years.	
X = Last Author	Last Author listed on the Citation	PM
X = Institution	Institutional Addresses listed on the Citation	PM
X = Key Word	Key Words listed on the Citation	PM
X = Journal	Journal listed on the Citation	PM
OPENNESS SHOCK CHARACTERISTICS		
Post NIH MoU	Dummy variable equal to 1 if Article j is associated with an openness MOU agreement (Cre-Lox, Onco) which is in effect in year t .	MGI
NIH MoU Window	Dummy variable equal to 1 if Article j is associated with an openness MoU agreement (Cre-Lox, Onco) which is in its initial period in year t .	MGI
MOUSE TECHNOLOGY CHARACTERISTICS (N = 4 technologies)		
Earliest Year	The publication year of the earliest mouse-article in the MGI database associated GM technology k .	MGI
Total Mice Created (1983 onward)	The total number of mice listed in the MGI database, with mouse-articles published from 1983 onward, associated with GM technology k .	MGI
MOUSE CREATION CHARACTERISTICS (N = 78 technology-year observations)		
Annual Mouse Creation	The number of mouse-articles published in Year t which introduce mice created using GM technology k .	MGI
New Creation Journals	Count of unique Journals publishing mouse-articles of GM technology k in Year t , which are “new” and have not published mouse-articles for GM technology k in prior years.	PM
Old Creation Journals	Count of unique Journals publishing mouse-articles of GM technology k in Year t , which are NOT “new” and <i>have</i> published mouse-articles for GM technology k in prior years.	PM

TABLE 2: SUMMARY STATISTICS

VARIABLE	N	MEAN	STD. DEV.	MIN	MAX
PUBLICATION CHARACTERISTICS (N = 2,171 original publications)					
Publication Year	2171	1995.58	2.34	1987	1998
# of Authors	2171	7.110	3.407	1	34
Total Forward Citations	2171	210.8	230.4	1	2543
CITATION-YEAR CHARACTERISTICS (N = 22,265 citation-year observations)					
Citation Year	22265	2001.19	3.26	1993	2006
Annual Citations	22265	19.41	21.67	0	336
High Quality Annual Citations	22265	4.313	5.868	0	71
CITATION-YEAR MARGIN CHARACTERISTICS (N = 22,265 citation-year observations)					
Basic Citations	22265	9.176	11.239	0	151
Applied Citations	22265	7.446	10.778	0	157
New Last-Authors	22265	11.708	13.479	0	243
Old Last-Authors	22265	3.941	5.349	0	58
New Institutions	22265	17.541	17.829	0	287
Old Institutions	22265	10.212	13.657	0	135
New Key Words	22265	74.878	67.277	0	794
Old Key Words	22265	55.371	61.115	0	620
New Journals	22265	7.911	7.842	0	94
Old Journals	22265	6.168	7.608	0	81
OPENNESS AGREEMENT CHARACTERISTICS (N = 22,265 citation-year observations)					
Post NIH MoU	22265	0.036	0.187	0	1
NIH MoU Window	22265	0.011	0.105	0	1
Post Cre-lox MoU	22265	0.009	0.093	0	1
Cre-lox MoU Window	22265	0.002	0.045	0	1
Post Onco MoU	22265	0.027	0.164	0	1
Onco MoU Window	22265	0.009	0.095	0	1
MOUSE CREATION CHARACTERISTICS (N = 78 technology-year observations)					
Annual Mouse Creation	78	108.192	171.586	0	875
New Creation Journals	78	6.987	5.805	0	24
Old Creation Journals	78	16.128	18.129	0	75

TABLE 3: SUMMARY STATISTICS BY MOUSE TECHNOLOGY

VARIABLE	MOUSE TECHNOLOGY			
	CRE-LOX	ONCO	KNOCK-OUT	SPONTANEOUS
PUBLICATION CHARACTERISTICS (N = 2,171 original publications)				
# of Original Publications	28	102	1895	146
Publication Year	1996.7	1993.4	1995.8	1993.5
# Authors	5.250	6.020	7.353	5.075
Total Citations	144.43	167.78	226.25	52.73
CITATION-YEAR CHARACTERISTICS (N = 22,265 citation-year observations)				
Annual Citations	14.719	12.297	21.285	3.955
MOUSE TECHNOLOGY CHARACTERISTICS (N = 4 technologies)				
Earliest Year	1992	1983	1989	1915
Total Mice Created (1983 onward)	1159	401	5980	911
MOUSE CREATION CHARACTERISTICS (N = 78 technology-year observations)				
Annual Mouse Creation	28.429	50.217	351.471	37.958
New Creation Journals	5.000	5.609	13.412	4.917
Old Creation Journals	7.286	14.783	37.765	7.250

TABLE 4: IMPACT OF OPENNESS ON FOLLOW-ON RESEARCH

	NEGATIVE BINOMIAL Dep Var = ANNUAL CITATIONS [Incident rate ratios reported in square brackets] Estimated coefficients in 2nd line. (Block bootstrapped SEs reported in parentheses)			
	(4-1) Baseline Model	(4-2) Baseline Model with Treatment Effect Dynamics	(4-3) Treatment Effects by Cre & Onco MoU	(4-4) Baseline Model, DV = Citations from High Quality Journals only [#]
Post NIH MoU	[1.302]*** 0.264 (0.062)			[1.409]*** 0.343 (0.080)
Post NIH MoU, Short-term ⁺⁺		[1.220]*** 0.199 (0.064)		
Post NIH MoU, Long-term ⁺⁺⁺		[1.429]*** 0.357 (0.074)		
Post Cre-lox MoU			[1.467]*** 0.383 (0.115)	
Post Onco MoU			[1.267]*** 0.236 (0.060)	
CONTROL VARIABLES				
- NIH MoU Window ⁺	[1.146]** 0.136 (0.065)	[1.149]** 0.139 (0.058)	-	[0.954] -0.047 (0.092)
- Cre MoU Window	-	-	[1.069] 0.067 (0.089)	-
- Onco MoU Window	-	-	[1.188]*** 0.172 (0.043)	-
Age FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
Article FEs	Yes	Yes	Yes	Yes
Log-likelihood	-55919.8	-55906.1	-55912.4	-34112.8
# of Observations	22265	22265	22265	21574

Significance levels: * 10% ** 5% *** 1%

+ Window is defined as the year of the MoU and the following year (for Cre-lox mice: 1998/1999; for Onco mice: 1999/2000)

++ Short-term is defined as the three years following the window after the MoU (for Cre-lox mice: 2000-2002; for Onco mice: 2001-2003).

+++ Long-term is defined as the years following the window and the short-term period after the MoU (for Cre-lox mice: 2003 onward; for Onco mice: 2004 onward).

[#] For this regression we use a modified dependent variable that captures only those annual citations that appear in a subset of high quality journals ranked by ISI impact factor.

Tests of Differences Between Coefficients:

(4-1): $\beta(\text{Post-NIH MoU}) - \beta(\text{NIH MoU Window})$:
Estimate = 0.129; SE = 0.033; Prob >|z| < 0.001

(4-2): $\beta(\text{Post-NIH MoU Long-term}) - \beta(\text{Post-NIH MoU Short-term})$:
Estimate = 0.158; SE = 0.040; Prob >|z| < 0.001

TABLE 5: IMPACT OF OPENNESS ON CITATIONS BY NEW VS OLD ‘LAST AUTHORS’ AND NEW VS. OLD INSTITUTIONS

	STACKED NEGATIVE BINOMIAL [Incident rate ratios reported in square brackets] Estimated coefficients in 2nd line. (Block bootstrapped SEs reported in parentheses)							
	(5-1a) DV= New Authors	(5-1b) DV= Old Authors	(5-2a) DV= New Authors	(5-2b) DV= Old Authors	(5-3a) DV= New Authors	(5-3b) DV= Old Authors	(5-3a) DV= New Institutions	(5-3b) DV= Old Institutions
Post NIH MoU	[1.379]*** 0.321 (0.065)	[1.135] 0.127 (0.088)					[1.269]*** 0.238 (0.052)	[1.127]* 0.120 (0.066)
Post NIH MoU, Short-term			[1.276]*** 0.244 (0.062)	[1.064] 0.062 (0.078)				
Post NIH MoU, Long-term			[1.537]*** 0.430 (0.071)	[1.224]*** 0.202 (0.073)				
Post Cre-lox MoU					[1.649]** 0.500 (0.203)	[1.189] 0.173 (0.211)		
Post Onco MoU					[1.305]*** 0.266 (0.076)	[1.160] 0.148 (0.108)		
CONTROL VARIABLES								
Window FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes ⁺		Yes ⁺		Yes ⁺		Yes ⁺	
Article FEs	Yes		Yes		Yes		Yes	
Log-likelihood	-86889.3		-86874.1		-86877.2		-114094.0	
# of Observations	42802		42802		42802		42830	

Significance levels: * 10% ** 5% *** 1%

+ Calendar year fixed effects include a set of indicator variables common to both margins in a given regression, and a linear difference variable which allows for a constant difference in growth rates between the two margins.

Tests of Differences Between Coefficients:

(5-1): $\beta(\text{Post-NIH MoU effect on New Authors}) - \beta(\text{Post-NIH MoU effect on Old Authors}):$
Estimate = 0.194; SE = 0.042; Prob >|z| < 0.001

(5-2): $\beta(\text{Post-NIH MoU Short-term effect on New Authors}) - \beta(\text{Post-NIH MoU Short-term effect on Old Authors}):$
Estimate = 0.181; SE = 0.047; Prob >|z| < 0.001
 $\beta(\text{Post-NIH MoU Long-term effect on New Authors}) - \beta(\text{Post-NIH MoU Long-term effect on Old Authors}):$
Estimate = 0.227; SE = 0.042; Prob >|z| < 0.001

(5-3): $\beta(\text{Post-Cre-lox MoU effect on New Authors}) - \beta(\text{Post-Cre-lox MoU effect on Old Authors}):$
Estimate = 0.327; SE = 0.064; Prob >|z| < 0.001
 $\beta(\text{Post-Onco MoU effect on New Authors}) - \beta(\text{Post-Onco MoU effect on Old Authors}):$
Estimate = 0.118; SE = 0.054; Prob >|z| = 0.029

(5-4): $\beta(\text{Post-NIH MoU effect on New Institutions}) - \beta(\text{Post-NIH MoU effect on Old Institutions}):$
Estimate = 0.118; SE = 0.035; Prob >|z| = 0.001

**TABLE 6: IMPACT OF OPENNESS ON CITATIONS
WITH NEW VS. OLD KEY WORDS**

	STACKED NEGATIVE BINOMIAL [Incident rate ratios reported in square brackets] Estimated coefficients in 2nd line. (Block bootstrapped SEs reported in parentheses)							
	(6-1a) DV=New Key Words	(6-1b) DV=Old Key Words	(6-2a) DV= New Key Words	(6-2b) DV= Old Key Words	(6-3a) DV=New Key Words	(6-3b) DV=Old Key Words	(6-4a) DV= New Journals	(6-4b) DV= Old Journals
Post NIH MoU	[1.260]*** 0.231 (0.070)	[0.925] -0.078 (0.075)					[1.381]*** 0.323 (0.076)	[1.201]** 0.183 (0.084)
Post NIH MoU, Short-term			[1.178]*** 0.164 (0.061)	[0.882]* -0.126 (0.066)				
Post NIH MoU, Long-term			[1.405]*** 0.340 (0.070)	[0.989] -0.011 (0.071)				
Post Cre-lox MoU					[1.399]* 0.336 (0.202)	[0.879] -0.129 (0.194)		
Post Onco MoU					[1.208]*** 0.189 (0.062)	[0.955] -0.046 (0.076)		
CONTROL VARIABLES								
Window FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes ⁺		Yes ⁺		Yes ⁺		Yes ⁺	
Article FEs	Yes		Yes		Yes		Yes	
Log-likelihood	-179179.1		-179162.5		-179146.0		-88007.3	
# of Observations	44488		44488		44488		42830	

Significance levels: * 10% ** 5% *** 1%

+ Calendar year fixed effects include a set of indicator variables common to both margins in a given regression, and a linear difference variable which allows for a constant difference in growth rates between the two margins.

Tests of Differences Between Coefficients:

(6-1): $\beta(\text{Post-NIH MoU effect on New Key Words}) - \beta(\text{Post-NIH MoU effect on Old Key Words}):$
Estimate = 0.310; SE = 0.038; Prob >|z| < 0.001

(6-2): $\beta(\text{Post-NIH MoU Short-term effect on New Key Words}) - \beta(\text{Post-NIH MoU Short-term effect on Old Key Words}):$
Estimate = 0.290; SE = 0.038; Prob >|z| < 0.001

$\beta(\text{Post-NIH MoU Long-term effect on New Key Words}) - \beta(\text{Post-NIH MoU Long-term effect on Old Key Words}):$
Estimate = 0.351; SE = 0.035; Prob >|z| < 0.001

(6-3): $\beta(\text{Post-Cre-lox MoU effect on New Key Words}) - \beta(\text{Post-Cre-lox MoU effect on Old Key Words}):$
Estimate = 0.466; SE = 0.059; Prob >|z| < 0.001

$\beta(\text{Post-Onco MoU effect on New Key Words}) - \beta(\text{Post-Onco MoU effect on Old Key Words}):$
Estimate = 0.235; SE = 0.039; Prob >|z| < 0.001

(6-4): $\beta(\text{Post-NIH MoU effect on New Journals}) - \beta(\text{Post-NIH MoU effect on Old Journals}):$
Estimate = 0.140; SE = 0.043; Prob >|z| = 0.001

TABLE 7: ROBUSTNESS TESTS FOR A PRE-SHOCK TREATMENT TREND FOR RESULTS ON OVERALL CITATIONS, NEW VS OLD AUTHORS AND NEW VS OLD KEYWORDS

	[Incident rate ratios reported in square brackets] Estimated coefficients in 2 nd line. (Block bootstrapped SEs reported in parentheses)				
	NEGATIVE BINOMIAL	STACKED NEGATIVE BINOMIAL			
	(7-1) DV= Annual Citations With Treatment Trends	(7-2a) DV= New Authors With Treatment Trends	(7-2b) DV= Old Authors With Treatment Trends	(7-3a) DV= New Keywords With Treatment Trends	(7-3b) DV= Old Keywords With Treatment Trends
Post NIH MoU	[1.145]* 0.135 (0.078)	[1.117] 0.111 (0.091)	[1.034] 0.033 (0.078)	[1.127] 0.120 (0.102)	[0.984] -0.016 (0.118)
Treatment Group Age Trend per Year	[1.003] 0.003 (0.015)	[1.014] 0.014 (0.018)	[1.000] -0.000 (0.020)	[1.001] 0.001 (0.025)	[0.997] -0.003 (0.029)
Post NIH MoU Change in Trend per Year	[1.050]*** 0.049 (0.017)	[1.052]** 0.051 (0.025)	[1.046]* 0.045 (0.025)	[1.053]** 0.052 (0.025)	[1.045] 0.044 (0.028)
CONTROL VARIABLES					
NIH MoU Window	[1.114]** 0.108 (0.047)	[1.079] 0.076 (0.062)	[1.091] 0.087 (0.071)	[1.081] 0.078 (0.068)	[0.860]** -0.151 (0.063)
Age FEs	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes ⁺		Yes ⁺	
Article FEs	Yes	Yes		Yes	
Log-likelihood	-55899.5	-86859.4		-179152.1	
# of Observations	22265	42802		44488	

Significance levels: * 10% ** 5% *** 1%

+ Calendar year fixed effects include a set of indicator variables common to both margins in a given regression, and a linear difference variable which allows for a constant difference in growth rates between the two margins.

Tests of Differences Between Coefficients:

(7-2): $\beta(\text{Post-NIH MoU effect on New Authors}) - \beta(\text{Post-NIH MoU effect on Old Authors}):$
Estimate = **0.078**; SE = **0.077**; Prob >|z| = **0.312**

(7-3): $\beta(\text{Post-NIH MoU effect on New Key Words}) - \beta(\text{Post-NIH MoU effect on Old Key Words}):$
Estimate = **0.277**; SE = **0.056**; Prob >|z| < **0.001**

**TABLE 8: IMPACT OF OPENNESS ON CITATIONS
IN BASIC VS. APPLIED JOURNALS**

	STACKED NEGATIVE BINOMIAL [Incident rate ratios reported in square brackets] Estimated coefficients in 2 nd line. (Block bootstrapped SEs reported in parentheses)			
	(8-1a) DV= Basic Journal Citations	(8-1b) DV= Applied Journal Citations	(8-2a) DV= Basic Journal Citations	(8-2b) DV= Applied Journal Citations
Post NIH MoU	[1.262]*** 0.233 (0.066)	[1.301]*** 0.263 (0.061)		
Post Cre-lox MoU			[2.212]*** 0.794 (0.126)	[1.073] 0.070 (0.105)
Post Onco MoU			[1.076] 0.073 (0.062)	[1.565]*** 0.448 (0.075)
CONTROL VARIABLES				
MoU Window FEs	Yes	Yes	Yes	Yes
Age FEs	Yes	Yes	Yes	Yes
Year FEs	Yes ⁺		Yes ⁺	
Article FEs	Yes		Yes	
Log-likelihood	-105989.0		-105894.7	
# of Observations	44530		44530	

Significance levels: * 10% ** 5% *** 1%

+ Calendar year fixed effects include a set of indicator variables common to both margins in a given regression, and a linear difference variable which allows for a constant difference in growth rates between the two margins.

Tests of Differences Between Coefficients:

(8-1): $\beta(\text{Post-NIH MoU effect on Basic Journal Citations}) - \beta(\text{Post-NIH MoU effect on Applied Journal Citations})$:
Estimate = -0.030; SE = 0.072; Prob >|z| = 0.676

(8-2): $\beta(\text{Post-Cre-lox MoU effect on Basic Journal Citations}) - \beta(\text{Post-Cre-lox MoU effect on Applied Journal Citations})$:
Estimate = 0.724; SE = 0.122; Prob >|z| < 0.001
 $\beta(\text{Post-Onco MoU effect on Basic Journal Citations}) - \beta(\text{Post-Onco MoU effect on Applied Journal Citations})$:
Estimate = -0.375; SE = 0.086; Prob >|z| < 0.001

TABLE 9: IMPACT OF OPENNESS ON OVERALL MOUSE CREATION AND MOUSE CREATION IN NEW VS. OLD JOURNALS

	[Incident rate ratios reported in square brackets] Estimated coefficients in 2 nd line. (Block bootstrapped SEs reported in parentheses)					
	NEGATIVE BINOMIAL		STACKED NEGATIVE BINOMIAL			
	(9-1) DV= Annual Mouse Creation	(9-2) DV= Annual Mouse Creation	(9-3a) DV= New Creation Journals	(9-3b) DV= Old Creation Journals	(9-4a) DV= New Creation Journals	(9-4b) DV= Old Creation Journals
Cre-lox MoU Window	[1.657]*** 0.505 (0.191)	[1.318] 0.276 (0.280)	[3.629]*** 1.289 (0.316)	[0.767]* -0.265 (0.135)	[3.367]*** 1.214 (0.396)	[0.797] -0.227 (0.422)
Post Cre-lox MoU	[1.545]* 0.435 (0.252)	[1.208] 0.189 (0.296)	[1.763] 0.567 (0.449)	[1.271]*** 0.240 (0.076)	[1.701] 0.531 (0.531)	[1.276] 0.244 (0.577)
Onco MoU Window	[1.579]* 0.457 (0.248)	[1.031] 0.031 (0.097)	[1.556]*** 0.442 (0.095)	[1.007] 0.007 (0.062)	[1.075] 0.072 (0.292)	[0.756] -0.280 (0.299)
Post Onco MoU	[1.835]*** 0.607 (0.113)	[0.972] -0.028 (0.182)	[2.586]*** 0.950 (0.125)	[1.256]*** 0.228 (0.077)	[1.462] 0.380 (0.429)	[0.841] -0.173 (0.440)
CONTROL VARIABLES						
Age Controls	Yes ⁺	Yes ⁺	Yes ⁺	Yes ⁺	Yes ⁺	Yes ⁺
Year FEs	Yes	Yes	Yes ⁺⁺		Yes ⁺⁺	
Technology FEs	Yes	Yes	Yes		Yes	
Log-likelihood	-287.449	-518.927	-335.559		-580.809	
# of Technologies	4	8	4		8	
# of Observations	78	136	156		272	

Significance levels: * 10% ** 5% *** 1%

+ Age controls include a cubic polynomial in the age of the technology, plus a linear correction in age for the much-older spontaneous technology. Note that there are separate effects for the two margins in stacked regressions.

++ Calendar year fixed effects include a set of indicator variables common to both margins in a given regression, and a linear difference variable which allows for a constant difference in growth rates between the two margins.

Tests of Differences Between Coefficients:

(9-3): $\beta(\text{Cre-lox MoU Window effect on New Creation Journals}) - \beta(\text{Cre-lox MoU Window effect on Old Creation Journals}):$

Estimate = **1.554**; *SE* = **0.451**; *Prob >|z|* = **0.001**

$\beta(\text{Onco MoU Window effect on New Creation Journals}) - \beta(\text{Onco MoU Window effect on Old Creation Journals}):$

Estimate = **0.435**; *SE* = **0.050**; *Prob >|z|* < **0.001**

$\beta(\text{Post-Onco MoU effect on New Creation Journals}) - \beta(\text{Post-Onco MoU effect on Old Creation Journals}):$

Estimate = **0.722**; *SE* = **0.078**; *Prob >|z|* < **0.001**

(9-4): $\beta(\text{Cre-lox MoU Window effect on New Creation Journals}) - \beta(\text{Cre-lox MoU Window effect on Old Creation Journals}):$

Estimate = **1.442**; *SE* = **0.195**; *Prob >|z|* < **0.001**

$\beta(\text{Onco MoU Window effect on New Creation Journals}) - \beta(\text{Onco MoU Window effect on Old Creation Journals}):$

Estimate = **0.352**; *SE* = **0.140**; *Prob >|z|* = **0.012**

$\beta(\text{Post-Onco MoU effect on New Creation Journals}) - \beta(\text{Post-Onco MoU effect on Old Creation Journals}):$

Estimate = **0.553**; *SE* = **0.241**; *Prob >|z|* = **0.022**