

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

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

May 1, 2009

Abstract

This paper argues that openness of upstream research does not simply encourage higher levels of downstream exploitation, it also 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. Moreover, this reflects increased exploration of more diverse research paths. Overall, our findings highlight a neglected cost of IP: reductions in the diversity of experimentation that follows from a single idea.

1 Introduction

While a whole recent literature examines the causal role of alternative institutional forms and policies, in particular Intellectual Property (IP) protection systems, on downstream R&D and innovation¹, there is far less research on the causal impact of the institutional and policy environment on the rate and direction of upstream research. However, the past three decades have seen a significant increase in the scope of formal intellectual property (IP) rights,

*MIT, Sloan School of Management

†Harvard University and NBER

‡ECARES, Université Libre de Bruxelles and CEPR

§Harvard University

¶Kellogg School of Management, Northwestern University and NBER

¹E.g. see Kortum and Lerner (2000).

such as patents, over knowledge traditionally maintained in the public domain.²Consequently, American universities currently receive over 3,000 U.S. patents each year and maintain a portfolio of over 40,000 patents.³ This dramatic expansion in IP rights over the earliest stages of research has caused widespread debate about the effectiveness of incentives for innovation.⁴ This debate is grounded in the notion that innovation is a step-by-step process in which discoveries generated in one stage serve as essential inputs into the next. The implications of expanding IP rights in the earliest stages of the innovation process are mixed. On the one hand, early-stage IP may be important to encourage the establishment of new research lines, since upstream researchers can thereby avoid expropriation by downstream researchers, as stressed in particular by Scotchmer (1996). On the other hand, by requiring downstream innovators to contend with a large number of fragmented upstream IP rights, their projects may suffer from "gridlock" as a result of transaction costs and complexity.⁵

By highlighting a single step-by-step research line, the current debate 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.⁶In 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.

What then is the role played by upstream IP rights when follow-on research includes both horizontal exploration as well as vertical exploitation? Interestingly, 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 (David and Dasgupta; 1994; David, 2003; Stern, 2004).

To guide our empirical analysis we use the multi-stage research framework developed by Aghion, Dewatripont and Stein (2007) who emphasize the role of freedom for researchers – defined as the granting of control rights allowing researchers to select their research direction. We then analyze the role of open-

²See Mowery, et al (2001) and Heller (2008).

³See Owen-Smith & Powell (2003).

⁴See Scotchmer (1991, 1996).

⁵See Heller and Eisenberg (1998) and Heller (2008).

⁶See Bresnahan and Trajtenberg (1995), and Rosenberg and Trajtenberg (2001).

ness in this framework. In this setting, openness not only impacts innovation incentives within a given research line but also encourages exploration and investment in new and speculative research directions. We identify three main channels whereby openness can influence the level and nature of scientific research. First, by reducing the costs of accessing key research inputs, openness encourages new researchers to enter, thus increasing the diversity of academic research participants.⁷Second, 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, our theoretical discussion suggests that, particularly in research settings characterized by high levels of freedom, openness not only increases the overall flow of research output, it should also be closely associated with the establishment and exploration of entirely new research lines. Moreover, while openness should affect both basic and applied research, the impact on basic research will, we predict, dominate when researchers in the pre-openness period face high fixed costs of initiating a new line of research. In 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 technology (Cre-Lox and Onco, respectively) 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 single 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

⁷Note 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 publicly subsidized research or through public buy-outs of (private) patents as in Kremer (1998).

follow-on research. First, each genetically engineered mouse is associated with a journal article that describes its initial development; as such, we are able to construct a sample based on "mouse-articles" affected by the NIH agreements and a control sample of mouse-articles unaffected by the agreements (based on Knock-out or Spontaneous technologies). Second, both the timing and the scope of the NIH-MoU 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 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 been linked to particular mouse-mouse), 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). Thus we develop three distinctive empirical tests that map to the three core claims of our theoretical framework.

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. These findings survive a whole set of robustness checks. Finally, Cre-Lox and Onco mice differed in whether researchers had any access prior to the NIH agreements (but faced some threat of IP enforcement) thus leading to differences in the likely impact of the NIH

MoUs. While the mice covered by the Onco agreement were available prior to the MoU, researchers were responsible for separately signing licenses as they moved to downstream applications. Mice based on the Cre-Lox technology were much more limited in their distribution. Reflecting these differences (and our theoretical predictions), mouse-articles associated with the Cre-Lox agreement experience a significant increase in citations by basic research journals, while mouse-articles associated with the Onco agreement realize a citation increase in applied research journals. Overall, these results are consistent with the view that the NIH agreements facilitate access to research inputs. As a result, in academic settings where control rights over research direction lie in the hands of researchers, increased openness has at least as large an effect on enhancing the scope and diversity of horizontal exploration as it does on inducing vertical exploitation along well-defined research lines.

The paper is organized as follows. Section 2 presents our theoretical framework and develops its main predictions concerning the effects of increased openness on the horizontal and vertical flow of research. Section 3 describes the experiment. Section 4 outlines our identification strategy. Section 5 presents the data and summary statistics. Section 6 presents the empirical results and Section 7 concludes.

2 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 researchers to provide access to her own ideas and share them as she sees fit. We shall argue that increased openness has three main effects on basic research. First, as noted in the introduction, openness tends to favor more applied research, possibly at the expense of more basic research, as it reduces the extent to which upstream researchers can appropriate the returns from their own research. Second, openness makes more likely that stage- $i + 1$ researchers will know about, and therefore build upon, the ideas of stage- i researchers, which in turn will increase the *ex ante* incentive to undertake stage i . Third, openness fosters more 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. We now discuss these various effects of openness, first abstracting from control rights considerations and focusing on the effects of openness on basic and applied research on a given line, then emphasizing the complementarity between openness and freedom and the resulting effect of openness on the diversity of lines.

2.1 Openness down a particular research line

Consider a two-stage research line. Each stage requires one researcher. Success at each stage, occurs with probability p , and moves research up to the next stage.

Then, as long as we focus on a single research line, a first effect of openness is that 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 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 .

Assume now that openness has an additional effect, by also increasing 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).⁸

To sum up, openness should be expected to foster downstream research thanks to higher appropriability. As for upstream research, the adverse effect of downstream appropriability can at times be outweighed by a probability of finding a good match interested in pursuing the research agenda.

2.2 Openness and diversification

In this subsection we enrich the above framework by introducing the notion of academic freedom, drawing on Aghion, Dewatripont and Stein (2008), henceforth ADS. We then analyze the interplay between freedom and openness, and in particular 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.

⁸If 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 $\frac{d\Pi_1}{d\psi} > 0$.

2.2.1 Introducing academic freedom: ADS

We keep assuming 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 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 that particular 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.⁹

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.

⁹In 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.

2.2.2 When is freedom optimal?

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 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 NPVs 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.

2.2.3 The value of experimentation

Note 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)¹⁰.

¹⁰See ADS for details.

2.2.4 Complementarity between openness and freedom: diversification effects

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 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 increased 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), for why increased openness should foster free research and therefore the creation of new lines, is that one particular feature of academic institutions which help them enforce the commitment not to monitor individual scientists' research agenda, is that they typically are non-profit institutions. This in turn makes them less willing or less able to incur the cost of monitoring researchers. But that 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 Testable predictions

Beyond the prediction that increasing the openness of critical early-stage research inputs should globally enhance the total flow of knowledge¹¹, and the prediction that the causal impact of a shift to greater openness should be to generate more research over the long-run (not simply a short-run boost)¹², the most important predictions from our model relate to the types of research and researchers most likely to be impacted by an “openness shock” in a world where researchers have control rights on their research activities.¹³ Three predictions stand out. 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 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.

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.¹⁴ 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

¹¹This prediction accords for example with a recent study estimating the significant and positive impact of Biological Resource Centers that make key research materials widely available to researchers (Furman and Stern, 2008).

¹²In other words, because the shift to greater openness is an enduring condition of key innovation inputs (under our model) and such inputs can be valuable to follow-on researchers over a long period - generating not one but multiple research lines - we would expect to see a long-run move to greater follow-on research, not simply a one time shock.

¹³In our particular empirical setting, the openness shock is focused directly and exclusively on academic (public-sector) researchers. We therefore do not make specific predictions regarding the overall balance of innovation between the public and the private sector.

¹⁴We refer the reader to Murray (2009) for more details and references on the history of intellectual property and openness in the mouse genetics community.

in mouse genetics relied on “spontaneous mutations” for their disease studies: researchers bred mice that naturally exhibited particular disease-linked symptoms or behaviors.¹⁵ 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.¹⁶ 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.¹⁷ Three breakthroughs were particularly important. First, in a discovery awarded the 2007 Nobel Prize in Medicine, Mario Capecchi of 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.¹⁸ In 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 en-

¹⁵Given 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)).

¹⁶See Rader (2004).

¹⁷The 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).

¹⁸The 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>).

gineered organisms and the Bayh-Dole Act affirmatively allowed universities to seek patent protection and licensing revenues from Federally-funded research.¹⁹ 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 pathway²⁰).

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, Oncomice and Cre-Lox mice – were covered by relatively broad patents.²¹ In 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

¹⁹These 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)

²⁰See Murray (2007).

²¹Knock-out mice were covered under U.S. Patent 4,687,737, Oncomice under U.S. Patent 4,736,866 and Cre-lox mice under U.S. Patent 4,959,317.

academics, and claimed that the lack of access to these mice significantly reduced their freedom to pursue their own research agendas.²² Individual 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 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 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.²³ 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).²⁴ 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.

²²As 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”.

²³In 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.

²⁴See Murray (2009).

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 Oncomice to researchers prior to the 1999 MoU.

Over a two-year period, life sciences researchers seeking to take advantage of 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.²⁵ 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 Oncomice 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.

²⁵We discuss this point in more detail in Section 6.2 below.

4 Empirical strategy

Our theoretical framework suggests that the level and nature of follow-on research depend not only upon the quality and type of research inputs available but also upon the degree of "openness" of these research inputs. To test this idea, 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 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.²⁶ 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).²⁷ 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 sci-

²⁶Our 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).

²⁷While 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.

entific 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 similar scientific quality and importance, differing only insofar as the patent over Knock-out technology was unenforced by the University of Utah.²⁸

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.

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.²⁹ 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:

²⁸We 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.

²⁹Consistent with our discussion in Section 3, the window period for the Cre-Lox period covers 1998 and 1999, and the window period for Oncomice covers 1999 and 2000.

$$\begin{aligned} \text{Annual Citations}_{jt} = & f(\varepsilon_{jt}; \gamma_j + \beta_t + \delta_{t-PubYear} \\ & + \Psi_0 \text{NIH MoU Window}_{jt} + \Psi_1 \text{Post NIH MoU}_{jt}), \end{aligned} \quad (1)$$

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 divide the citations from each citation year 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 follow-on publications by authors who have previously cited a particular mouse-article (*Old Author Citations_{jt}*) versus those who have not previously cited a particular mouse-article (*New Author Citations_{jt}*).

More specifically, we jointly estimate the two equations:

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

and

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

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 $\delta_{t-PubYear}^{NEW}$ and $\delta_{t-PubYear}^{OLD}$ 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 $\Psi_1^{NEW} > \Psi_1^{OLD}$. 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 citations 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 define the key word from a given citation to be new if the key word has never been used in citations to a particular mouse-article in any prior year. For all key words that have appeared in citations to the particular mouse-article in prior years, we code this as old. 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 any 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.³⁰ 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.³¹

³⁰Our 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.

³¹It is worth noting that we do not examine the impact of openness on the academic/industry

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 from 1993 through 2006, the average *Citation Year_{jt}* is 2001. We also create an alternative dependent variable *High Quality Citations_{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:³²

- i. *New Last Author Citations_{jt}* and *Old Last Author Citations_{jt}*, with mean values equal to 11.7 and 3.9 respectively.
- ii. *New Institution Citations_{jt}* and *Old Institution Citations_{jt}*, with mean values equal to 17.5 and 10.2 respectively.
- iii. *New Keyword Citations_{jt}* and *Old Keyword Citations_{jt}*, with mean values equal to 74.9 and 55.4 respectively.
- iv. *New Journal Citations_{jt}* and *Old Journal Citations_{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

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.

³²Note 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.

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: $Cre\text{-}Lox\ MoU\ Window_{jt}$, $Post\ Cre\text{-}Lox\ MoU_{jt}$, $Onco\ MoU\ Window_{jt}$ and $Post\ Onco\ MoU_{jt}$.

To examine the short-term versus long-term impact of the NIH MoU, we also define a treatment variable, $Post\ NIH\ MoU, Short\text{-}Term$ equal to one for the first three years after the window period, and a separate measure, $Post\ NIH\ MoU, Long\text{-}Term$, 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 in all our analyses 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 the incidence-rate ratios (IRR). 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 our 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 Author Citations* and *Old Author Citations*. 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 our hypothesis: 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 Word Citations* and *Old Key Word Citations* 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 Journal Citations* and *Old Journal Citations*, 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

Finally 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

7 Conclusion

In this paper we argued that openness of upstream research does not simply encourage higher levels of downstream exploitation: it also 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. Not only did they boost the overall flow of follow-on research using specific engineered research mice, they also expanded the diversity of researchers working on particular research lines, and expanded the diversity of the research taking place. While both basic and applied research was significantly increased after the MoUs, given the tight access restrictions in addition to the appropriability concerns associated with Cre-lox mice, the Cre-lox MoU had a particularly striking impact on basic research. Of course, our interpretation depends upon the extent to which these MoUs were truly exogenous. While they certainly reflected the endogenous choice of DuPont, JAX, and the NIH, there is strong evidence to suggest that the timing of these MoUs was unanticipated. Indeed the fact that the long-term effect is significantly higher than the short-term effect and that there is no pre-deposit trend provides empirical evidence in this regard. Our results therefore highlight a key limitation in the current literature on intellectual property and innovation - the potential restrictions intellectual property right may place on the diversity of research and researchers who would otherwise take a single powerful idea and experiment across multiple research lines.

The theory and empirical 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 particularly at very early stages of the research line, then may stifle exploratory projects that are key to diverse follow-on innovation. Second, our framework suggests that more attention be paid by economists to recent corporate attempts to generate new sources of profit 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. However, the less effective experience of DuPont and other companies that kept on enforcing patents while also attempting to engage with the open scientific community³³ suggests that the systematic analysis of the forces and trade-offs at work in an economic environment with both proprietary and open firms competing with each other and cooperating with open communities, merits future research.

³³See 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), "Towards 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] 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.
- [9] 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.
- [10] 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 .
- [11] Heller, M (2008), *Gridlock Economy: How Too Much Ownership Wrecks Markets, Stops Innovation and Costs Lives*, Philadelphia PA: Basic Books.
- [12] Heller, M, and Eisenberg, R (1998), "Can Patents Deter Innovation? The Anticommons in Biomedical Research", *Science*, 280, 698-701.
- [13] 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.
- [14] Howitt, P (2003), "The Economics of Science and the Future of Universities", The Timlin Lecture 2000, University of Saskatchewan.

- [15] 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*.
- [16] Kremer, M (1998), "Patent Buyouts: A Mechanism for Encouraging Innovation ", *The Quarterly Journal of Economics*, 113, 1137-1167.
- [17] Kortum, S, and Lerner, J (2000), "Assessing the Contribution of Venture Capital to Innovation ", *The RAND Journal of Economics*, 31, 674-692.
- [18] Lim, K (2004), "The Relationship Between Research and Innovation in the Semi-Conductor and Pharmaceutical Industries", *Research Policy*, (33), 287-321.
- [19] Merton, R (1973), *The Sociology of Science: Theoretical and Empirical Investigations*, Edited by Norman Storer. Chicago: University of Chicago Press.
- [20] Mokyr, J (2002), *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton: Princeton University Press.
- [21] 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.
- [22] 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.
- [23] Murray, F (2002), "Innovation as Co-Evolution of Scientific and Technological Networks: Exploring Tissue Engineering", *Research Policy*, 31, Issues 8-9, 1389-1403.
- [24] Murray, F (2007), "The Stem Cell Market: Patents & the Pursuit of Scientific Progress", *New England Journal of Medicine*, 356 (23), 2341-2343.
- [25] 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.
- [26] Nelson, R (1959), "The Simple Economics of Basic Scientific Research," *Journal of Political Economy* 67(3), 297-306.
- [27] O'Mahony, S (2003), "Guarding the Commons: How Community Managed Projects Protect Their Work" , *Research Policy*, 32, 1179-1198.
- [28] Rader, K (2004), *Making Mice*, Princeton, NJ: Princeton University Press.

- [29] 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.
- [30] 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>.
- [31] Scotchmer, S (1991), "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law", *Journal of Economic Perspectives*, 5, 29-41.
- [32] Scotchmer, S (1996), "Protecting Early Innovators: Should Second-Generation Products be Patentable?", *The Rand Journal of Economics*, 27, Summer 1996, 322-331.
- [33] Stern, S (2004), "Do Scientists Pay to Be Scientists", *Management Science*, 50, 835-853.
- [34] Tapscott, D, and Williams, A (2006), *Wikinomics: How Mass Collaboration Changes Everything*, New York, NY: Portfolio Press.

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 Citations	Count of Forward Citations to Article j in Year t where characteristic X of the citations is “new” and has not appeared in the citations to Article j in prior years.	
Old X Citations	Count of Forward Citations to Article j in Year t where characteristic X of the citations is NOT “new” and <i>has</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

TABLE 2: SUMMARY STATISTICS

VARIABLE	N	MEAN	STD. DEV.	MIN	MAX
PUBLICATION CHARACTERISTICS (N = 2,171 original publication)					
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-Author Citations	22265	11.708	13.479	0	243
Old Last-Author Citations	22265	3.941	5.349	0	58
New Institution Citations	22265	17.541	17.829	0	287
Old Institution Citations	22265	10.212	13.657	0	135
New Key Word Citations	22265	74.878	67.277	0	794
Old Key Word Citations	22265	55.371	61.115	0	620
New Journal Citations	22265	7.911	7.842	0	94
Old Journals Citations	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

TABLE 3: SUMMARY STATISTICS BY MOUSE TECHNOLOGY

VARIABLE	MOUSE TECHNOLOGY			
	CRE-LOX	ONCO	KNOCK-OUT	SPONTANEOUS
PUBLICATION CHARACTERISTICS (N = 2,171 original publication)				
# 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

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 sub-set 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 Author Citations	(5-1b) DV= Old Author Citations	(5-2a) DV= New Author Citations	(5-2b) DV= Old Author Citations	(5-3a) DV= New Author Citations	(5-3b) DV= Old Author Citations	(5-3a) DV= New Institution Citations	(5-3b) DV= Old Institution Citations
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): $\beta(\text{Post-NIH MoU effect on New Author Citations}) - \beta(\text{Post-NIH MoU effect on Old Author Citations}):$
Estimate = 0.194; SE = 0.042; Prob >|z| < 0.001

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

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

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

TABLE 6: IMPACT OF OPENNESS SHOCKS 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 Word Citations	(6-1b) DV=Old Key Word Citations	(6-2a) DV= New Key Word Citations	(6-2b) DV= Old Key Word Citations	(6-3a) DV=New Key Word Citations	(6-3b) DV=Old Key Word Citations	(6-4a) DV= New Journal Citations	(6-4b) DV= Old Journal Citations
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): $\beta(\text{Post-NIH MoU effect on New Key Word Citations}) - \beta(\text{Post-NIH MoU effect on Old Key Word Citations})$:
Estimate = 0.310; SE = 0.038; Prob >|z| < 0.001

(6-2): $\beta(\text{Post-NIH MoU Short-term effect on New Key Word Citations}) - \beta(\text{Post-NIH MoU Short-term effect on Old Key Word Citations})$:
Estimate = 0.290; SE = 0.038; Prob >|z| < 0.001
 $\beta(\text{Post-NIH MoU Long-term effect on New Key Word Citations}) - \beta(\text{Post-NIH MoU Long-term effect on Old Key Word Citations})$:
Estimate = 0.351; SE = 0.035; Prob >|z| < 0.001

(6-3): $\beta(\text{Post-Cre-lox MoU effect on New Key Word Citations}) - \beta(\text{Post-Cre-lox MoU effect on Old Key Word Citations})$:
Estimate = 0.466; SE = 0.059; Prob >|z| < 0.001
 $\beta(\text{Post-Onco MoU effect on New Key Word Citations}) - \beta(\text{Post-Onco MoU effect on Old Key Word Citations})$:
Estimate = 0.235; SE = 0.039; Prob >|z| < 0.001

(6-4): $\beta(\text{Post-NIH MoU effect on New Journal Citations}) - \beta(\text{Post-NIH MoU effect on Old Journal Citations})$:
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

	STACKED NEGATIVE BINOMIAL [Incident rate ratios reported in square brackets] Estimated coefficients in 2 nd line. (Block bootstrapped SEs reported in parentheses)				
	(7-1) DV= Annual Citations With Treatment Trends	(7-2a) DV= New Author Citations With Treatment Trends	(7-2b) DV= Old Author Citations With Treatment Trends	(7-3a) DV= New Keyword Citations With Treatment Trends	(7-3b) DV= Old Keyword Citations 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): $\beta(\text{Post-NIH MoU effect on New Author Citations}) - \beta(\text{Post-NIH MoU effect on Old Author Citations}):$
Estimate = 0.078; SE = 0.077; Prob >|z| = 0.312

(7-3): $\beta(\text{Post-NIH MoU effect on New Key Word Citations}) - \beta(\text{Post-NIH MoU effect on Old Key Word Citations}):$
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