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

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

December 24, 2008

Abstract

Scientific freedom and openness are hallmarks of academia: relative to their counterparts in industry, academics maintain discretion over their research agenda and allow others to build on their discoveries. This paper examines the relationship between openness and freedom, building on recent models emphasizing that, from an economic perspective, freedom is the granting of control rights to researchers. Within this framework, 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. In other words, within academia, restrictions on scientific openness (such as those created by formal intellectual property (IP)) may limit the diversity and experimentation of basic research itself. We test 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 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 increase is driven by a substantial increase in the rate of 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.

*MIT, Sloan School of Management
†Harvard University and NBER
‡ECARES, Universite Libre de Bruxelles and CEPR
§Harvard University
¶Kellogg School of Management, Northwestern University and NBER

1 Introduction

The past three decades have seen a significant increase in the scope of formal intellectual property (IP) rights, such as patents, over knowledge traditionally maintained in the public domain (Mowery, et al 2001; Heller 2008). American universities are granted over 3,000 U.S. patents per year and maintain a portfolio of over 40,000 patents (Owen-Smith & Powell 2003). Notably, nearly 25% of elite academic life sciences researchers hold at least one patent (Ding, Murray & Stuart 2006), mostly for discoveries arising from their university-based research (Azoulay, Ding & Stuart 2007). This dramatic expansion in property rights over the earliest stages of research and over key research inputs has caused widespread debate. In particular, it has shifted the economic analysis of patents away from traditional concerns over the costs of monopoly pricing in product markets (Nordhaus 1969, Scherer 1972) towards a focus on innovators incentives when property rights are shared across innovators, each of whom is working at a different stage of the knowledge production process (Scotchmer 1991, 1996; Aghion, Harris & Vickers 2000; Acemoglu & Akcigit 2006). This multi-stage characterization describes innovation as a step-by-step process in which discoveries generated in one stage serve as essential inputs into the next. In terms of realizing the value from a given research line, early-stage IP rights may be important factors encouraging the establishment of new research lines, since upstream researchers can subsequently offer incentives for research further along the line through appropriate contract design (Scotchmer 1996). At the same time, recent debates over the proliferation of upstream IP suggest that by requiring downstream innovators to contend with a large number of fragmented IP rights, their projects may suffer from "gridlock" as a result of transaction costs and complexity (Heller & Eisenberg 1998; Heller 2008).

Research highlighting a single step-by-step research line 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 (Breshnahan & Trajtenberg 1995; Romer 1990; Rosenberg & Trajtenberg 2001). In other words, ideas are non-rivalrous. Second, it may be extremely difficult in advance to precisely articulate the diversity and range of applications arising from a given upstream idea (Rosenberg 1996). Different individuals may have very different perceptions regarding the main application of an idea or the follow-on research projects they would prefer to pursue (Shane 2001). In other words, rather than focusing exclusively on the value generated along a single line, we argue that it may also be useful to consider whether multiple researchers are able to pursue a diverse range of "horizontal " follow-on experiments each of which may itself 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 regarding IP rights (or conversely openness) has focused 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. This paper builds on

recent research analyzing the distinctive incentives and control rights provided by the institutional regimes in academia versus industrial researchers (David and Dasgupta 1994, David 2001ab, David 2003), and more specifically on Aghion, Dewatripont & Stein 2007 who emphasize the role of academic freedom defined as the granting of control rights to researchers (see also Stern 2004). In particular, a very distinctive aspect of academia as opposed to industrial research is that academic researchers are free to establish new research lines, based on their perception of opportunities or on their pure curiosity-driven choices.

In this paper we extend the control-rights framework to 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. Second, relative to what would happen in the case of industrial research, openness makes free (academic) researchers more likely to engage in experiments that broaden the number and diversity of research lines, in part because subsequent openness implies that their research can itself have subsequent impact. Finally, there is of course a direct expropriation effect – an increase in the level of openness of an upstream research tool should encourage the exploitation of that tool in research which is already well down the research line and in the more applied phase. Overall, our theoretical discussion suggests that, particularly in free (academic) research, openness may not only increase the overall flow of research output, and in particular be closely associated with the establishment and exploration of entirely new research lines. Moreover, while there should be an effect of openness on both basic and applied research, the impact on basic research is more likely to dominate when researchers in the pre-openness period face high fixed costs of initiating a new line of research, while the applied research boost 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 and publish research using particular types of genetically engineered mice that were covered under two different patents (Cre-Lox mice and Onco mice, respectively). While DuPont had adopted stringent restrictions on licensing the mice for academic research prior to the MoUs, the agreements lifted these restrictions by implementing a simple contract, providing a royalty-free and costless license that specifically 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. As a result of these MoUs significantly enhancing the openness regarding these research tools, hundreds of varieties of Cre-lox or Onco mice that had been developed in the early 1990s suddenly became widely accessible to the academic research community.

Our empirical approach takes advantage of key aspects of our empirical

setting to develop and implement a differences-in-differences estimate of the impact of the NIH-MoU openness experiment on both the level and nature of follow-on research. First, each genetically engineered mouse is associated with a journal article that describes its initial development; as such, we are able to construct samples based on research articles that were affected or unaffected by the NIH agreements. Second, both the timing and the scope of the NIH-MoU were effectively unanticipated by the mouse genetics community, and so there was a fairly unexpected and dramatic shift in the level of openness in a reasonably short period of time. Finally, we are able to take advantage of detailed bibliometric data for articles citing the articles in either the treatment or control groups to characterize how the change in openness 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 published mouse-articles, approximately 10% of which experienced a shift in the level of scientific openness as the result of the NIH agreements. By comparing citations to the mouse-articles before and after the agreement (and comparing to the evolution of citations of 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 nature of citations differs after the shift in openness. Specifically, we construct measures capturing whether there is a shift in the size of the research community using a particular mouse (such as the number of new authors citing the mouse-paper), whether research is associated with the establishment of new research lines that had not previously used a particular mouse (such as whether the citations include keywords that had never been linked to particular mouse-paper), and whether the research is more basic versus applied (as captured by the journal in which the research is published). Thus we develop three distinctive empirical tests that map to the three claims of our core 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 increase in citations arises from articles that are published by "new" researchers or institutions. In other words, most of the incremental citations to a given mouse-article are by researchers working at institutions that had not cited that mouse-article prior to the NIH agreement. Next, our results offer direct evidence that scientific openness seems to be associated with the establishment of entirely new research lines: more specifically, increased openness leads 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 underlying research contributions of the citing articles. Finally, the two agreements – Cre-Lox and Oncomouse – differed in terms of whether researchers had access to the mice prior to the agreement at all (but faced some threat of IP enforcement). While the mice covered by the Oncomouse agreement

were available but 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 Oncomouse agreement realize also an increase in citation by applied research journals. Taken together, this evidence is consistent with the view that the NIH agreement facilitated access to research inputs, and that, at least in the academic setting where control rights over research direction is in the hands of researchers, the effects of openness have at least as large an effect on enhancing the scope and diversity of horizontal exploration as on inducing vertical exploitation along well-defined research lines.

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

2 Openness in scientific knowledge production

2.1 The value of academic freedom

In recent work, Aghion, Dewatripont and Stein (2007) (ADS) have argued that the allocation of control rights is central to knowledge production and innovation. In a simple multi-stage representation of the discovery process, they suggest that freedom is more important for the production of basic -or early stage- research compared to applied research. Their core idea is that in earlier stages of the research process, when monetary returns from the research line remain remote, it is optimal to leave control rights for agenda setting with the researcher. In other words, to promote academic freedom. In contrast, later stages in the research process it becomes optimal to have control rights over the research agenda be retained by the firm or lab.

Specifically, ADS consider research as multi-stage lines where the development of an economically valuable product starts with an initial idea I_0. If stage 1 is successful, there is a refined idea I_1; this refined idea can be further worked on to potentially generate an even-more-refined idea I_2, etc. There are a total of k stages after the initial idea. If and only if all k stages are successful, there is a final idea I_k which generates a marketable product with value V.

Suppose for simplicity that at any given stage it is optimal to hire a single researcher.[1] Assume that this researcher obtains a probability of success equal to $p < 1$ at any stage if he follows the success-maximizing (which we call "practical") research strategy at that stage. Assume however that, instead of the

[1] See ADS for an extension to the case with more than one researcher per stage.

practical strategy, a researcher may choose to follow an "alternative" strategy in working with an existing idea. Assume first, for the sake of the argument, that in this case the scientist has a zero individual probability of success. The interpretation is that the alternative strategy amounts to the scientist working on an activity that he enjoys more but that does not pay off in monetary terms (see the end of this section for another interpretation where 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 R. After being hired at stage j, the scientist is exposed to idea I_{j-1}, and then learns whether he would better enjoy following the practical strategy or the alternative strategy. If he is able to undertake his favored strategy, he suffers no disutility from working. However, if 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. the alternative strategy is ex ante non-contractible.[2]

Academia differs from private-sector research in that it leaves control rights over the choice of research strategy in the hands of the researcher. Thus if a research line is pursued in Academia, the researcher is paid wage $w_a = R$, 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 $w_p = R+(1-\alpha)z$ in order to work in the private sector. The $(1 - \alpha)z$ markup over the academic wage represents compensation for loss of creative freedom–the fact that scientists now always have to adopt the practical strategy, whether this turns out to coincide with their preferences or not.

2.1.1 When is academic freedom optimal?

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 $E(\pi_2^a) = \alpha pV - w_a$. This means that the private sector will yield a higher payoff than academia if and only if $(1 - \alpha)pV > (w_p - w_a)$, or equivalently $pV > z$.

[2] In 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.

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 $E(\pi_1^a) = \alpha p\Pi_2 - w_a$. The private sector will yield a higher payoff than academia at stage 1 if and only if $p\Pi_2 > z$.

Since $\Pi_2 < V$, it follows that if the private sector is value-maximizing at stage 1, it is also value-maximizing at stage 2. In particular it cannot be value maximizing to have academia operate at later stages than the private sector. The key result is therefore that academia will be the optimal governance structure at earlier stages and private sector research will be optimal at later stages. 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:

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

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.

2.1.2 Valuable experimentation

The ADS framework thus provides a rationale for 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 can lead to an entirely new research line, consistently with the idea that scientific discoveries do not follow a purely "linear" model. This does not alter the relative optimality of academia (resp. private research) in earlier (resp. later) stages of research. But it raises the desirability of academia, 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).

2.2 The main effects of openness

Now, let us introduce the idea of openness into the framework, where openness is broadly defined as any event or device that reduces a researcher's difficulty to access other researchers' ideas or 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, 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. This is the appropriability effect pointed out in the introduction. Second, openness makes it easier for stage-i researchers to "sell" their ideas to stage-$i+1$ researchers, which in turn encourages them 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 free researchers 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.

2.2.1 Within a line: facilitating downstream transmission

For simplicity, consider a two-stage research line where stages 1 and 2 are managed in academia. Suppose first that openness increases the extent to which stage 2 can extract rents from stage 1. Thus,

$$\Pi_2 = \alpha p V + \psi - w_a,$$

where ψ is the additional rent openness gives stage 2 at the expense of stage 1. The stage-1 value of the line can then be written as:

$$\Pi_1 = \alpha p(\Pi_2 - 2\psi) - w_a = \alpha^2 p^2 V - \alpha p \psi - (1 + \alpha p) w_a.$$

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:

$$\Pi_1 = \alpha p A(\psi)(\Pi_2 - 2\psi) - w_a = A(\psi)(\alpha^2 p^2 V - \alpha p \psi) - (1 + \alpha p) w_a.$$

In turn, this implies:

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

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).

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

2.2.2 Complementarity between openness and freedom: diversification effects

That more openness should also foster the creation of new lines in academia, follows from an additional consideration, which is 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, 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 increases the net present value of a research line operated in academia in a given stage i, from:

$$\alpha p \Pi_i - w_a$$

to:

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

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 do conduct these investigations.

Remark: That openness should enhance basic research and the creation of new lines, also implies that it should have a long-lasting effect on the flow of subsequent publications: the reason is that new lines take more time before maturing. Indeed, starting a new line means a positive probability of a potentially long dynamic flow of new discoveries until one potentially reaches the end of this line**.**

2.3 Testable predictions

The above discussion suggests that providing greater openness of critical inputs for follow-on innovation should enhance the total flow of knowledge building on materials that have become more open and accessible. This prediction is, of course, very intuitive, and accords with a recent study estimating the positive impact of Biological Resource Centers in making key research materials available to researchers (Furman & Stern 2008). Several more specific predictions emerge from our model, particularly for conditions when researchers enjoy freedom (such as academia). First, the causal impact of a shift to greater openness should be to generate more long-term researcher. In other words, because the shift to greater openness is an enduring condition of key innovation inputs (under our model) and such inputs can be of value 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. Second, and perhaps the most important predictions to be derived from our model, are those relating 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.[3]Four 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 openness and/or when control rights considerations are not first order, then we would expect the boost to arise in basic research with horizontal exploration dominating.

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

The remainder of this paper tests these ideas by taking advantage of two "natural experiments" that significantly shifted the level of openness associated with two types of genetically engineered mice, both crucial inputs into cumulative research in the modern life sciences. [4] To understand how we take advantage 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 (Boguski, 2002). Throughout the twentieth century, scientists in mouse genetics relied on "spontaneous mutations" for their disease studies: researchers bred mice that naturally exhibited particular disease-linked symptoms or behaviors. [5] 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 (referred to as "JAX" mice) (Rader 2004).

In the early 1980s, advances in molecular biology and the ability to manipulate embryonic stem cells (Evans et al. 1984) 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 (Ruddle et al. 1980, Brinster et al. 1981, Constantini & Lacy

[3] Given that in our particular empirical setting, the openness shock is focused directly and exclusively on public-sector researchers, we do not make specific predictions regarding the overall balance of innovation between the public and the private sector.

[4]This section draws on Murray (2009) which offers an analytical narrative history of the role of intellectual property and openness in the mouse genetics community.

[5]Given the value of such mutations, researchers also developed techniques significantly increasing the rate of mutation of research mice such as the exposing pregnant mice to high levels of radiation (Green & Roderick, 1966).

1981, Wagner et al. 1981ab).[6] 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 (Doetschman et al. 1987; Thomas & Capecchi 1987). Second, with partial funding from DuPont Corporation, Phil Leder and Tim Stewart at Harvard University developed the Oncomouse methods, which provided a means for inserting (rather than deleting) genes into an embryo, and therefore making mice susceptible to particular forms of cancer and other diseases (Stewart et al. 1984). Finally, researchers at the life sciences division of Du Pont Corporation developed the Cre-Lox technology - a precise "cutting and pasting" tool that turns "off" particular genes in specific tissue or organs (Sauer et al. 1987). By offering general-purpose tools to engineer discrete changes in the genetic profile of research mice, each of these 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.[7]

The revolution in mouse genetics occurred alongside several important shifts in the role of formal intellectual property in life sciences research. In 1980, the Supreme Court decision in *Diamond v Chakrabarty* established the patentability of genetically engineered organisms and the Bayh-Dole Act affirmatively allowed universities to seek patent protection and licensing revenues from Federally-funded research (Mowery et al 2004).[8] By the mid-1990s, US universities receiving over 3,000 patents each year. While many observers took this as an indicator universities' evolving role as engines of innovation and commercialization (Henderson, Jaffe & Trajtenberg 1998), some argued that strong IP rights over scientific research discoveries was detrimental to research productivity and effective cumulative discovery (Heller & 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 (Murray 2007))

[6]The 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 2003, Murray, 2009).

[7]The 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).

[8]These legal and policy shifts reflected, in part, increasing appreciation that certain types of university research were increasingly dual-natured: fundamental scientific discoveries could simultaneously have a very high degree of commercial utility (Stokes 1987; Murray & Stern 2007)

while other universities were accused of rent-seeking when they sought to enforce intellectual property claims over independent commercial discoveries (e.g. the University of Rochester's enforcement of its patents on the Cox-2 pathway (Shane & Somaya 2007).

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 – Knock-out mice, Oncomice and Cre-Lox mice – were covered under relatively broad patents.[9] 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 (i.e., these mice were distributed as JAX mice). 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 was able to acquire the 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 and founders) who wanted "freedom to operate" were obliged to obtain a license from DuPont when they sought to receive or share an Onco or Cre-Lox mouse. The detailed licensing agreement required annual disclosure to DuPont regarding experimental progress, limits on informal mouse exchange among academic researchers, and "reach through" rights allowing DuPont to automatically receive licensing revenue from any commercial applications developed using either Cre-Lox or Onco technology.

These limits to openness caused widespread discontent among the academic community. Academic researchers objected to the exercise of patent rights by a for-profit company as a significant limitation on the norms of openness among academics, and claimed that the lack of access to these mice significantly reduced their freedom to pursue their own research agendas (Murray, 2009). [10] 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.[11] While some researchers took advantage of informal sharing or access of Onco-mice from the JAX (opening themselves to a potential

[9] Knock-out mice were covered under U.S. Patent 4,687737, Oncomice under U.S. Patent 4,736,866 and Cre-lox mice under U.S. Patent 4959317.

[10] As 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" (Rajewsky quoted in Jaffe 2004).

[11] Paigan Quote

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 an open-access depository such as JAX.

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.[12] 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; it was also possible in principle to sign the DuPont licensing agreement, though very few institutions or researchers signed an actual agreement prior to 1999. Third, if a Cre-lox mouse was preferred, it might be available, but only 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) (Murray 2009). 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).

The degree of openness associated with both Cre-Lox and Oncomice mice shifted dramatically in 1998 and 1999, respectively. In response to considerable pressure from the academic community throughout the 1990s, the National Institutes of Health (NIH), with the direct involvement of NIH Direct and Nobel Laureate Harold Varmus, successfully negotiated two Memorandum of Understanding (MoU) among DuPont, the Jackson Laboratories (JAX), and the NIH. Together, they 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.

[12] In 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.

Over a two-year period, life sciences researchers seeking to take advantage of the revolution in mouse genetics thus experienced a significant shift in their ability to access and exploit research mice covered under these agreements, while experiencing no shift in the degree of openness associated with either Knock-out or Spontaneous mice. These differences provide the key source of variation that we exploit in our empirical work. 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 technology faced significant open-access restrictions until the time of the NIH agreements. Second, though the academic community lobbied continuously for increased openness regarding these research tools, both the timing of the agreement as well as its scope (essentially removing the main hurdles associated with access) were unanticipated (Marshall 1999). It is unlikely that researchers delayed projects in anticipation of such a comprehensive agreement; instead, researchers deterred by the DuPont licensing restrictions undertook different research projects or devoted themselves to other research directions (Murray 2009). Third, though the agreements cover a small number of DuPont-controlled patents, they impacted a large number of different 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, we can take advantage of the fact that these mice were developed and disclosed to the scientific literature at different times and that their follow-on use by other scientists was captured through the citation of these articles in follow-on scientific articles, to precisely identify the impact of the NIH openness agreements on the use of genetically engineered mice in follow-on scientific research.

Taken together, we believe that the openness shift associated with the NIH agreements accords well with the comparative statics developed in Section II. Specifically, engineered research mice are general purpose research tools that can be used in multiple research lines and at multiple research stages. Restricted access to these research tools has the potential to significantly impact both the horizontal and vertical research incentives and productivity. On the one hand, DuPont's patent enforcement strategy is a strong candidate for vertical impact, as the threat of reach-through rights for DuPont limits the incentives to pursue more applied research stages. At the same time, the complicated and costly process of obtaining "freedom to operate" is also likely to reduce the degree of horizontal exploration. As predicted by our theory, the restrictions and transaction costs imposed by DuPont's enforcement choices would limit investment in early-stage academic research that depends upon these research tools but where it is also difficult to anticipate the precise research direction, requirements or outputs.

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 cumulative 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 what the rate of follow-on research would have been if 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 towards capturing the essence of such an approach is to take advantage of natural institutional variation that involves shifts of 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) specialized research mice.[13] Our empirical strategy exploits several distinctive elements of the system by which scientific research is disclosed and cited. First, in most cases, new specialized research mice are disclosed through publication in scientific articles that describe their production and analyze distinctive features (we refer to these disclosures as mouse-articles). Notably, we are able to identify mouse-article pairs both for mice that are affected by the NIH agreements (i.e., Cre-Lox and Onco mouse-articles) and for mice unaffected by the NIH agreements (i.e., Knock-Out and spontaneous mouse articles).[14] 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

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

[14] While 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.

being an imperfect and noisy indicator of overall scientific impact, citations offer a systematic reflection of the process by which researchers acknowledge how their efforts at any one research stage build on the tools and knowledge developed by prior researchers (Hagstrom, 1965; Merton 1973; de Solla Price, 1976; Garfield 1979; Cole 2000). More specifically, our approach exclusively examines the citation patterns associated with mouse-articles. Our qualitative research indeed suggests that nearly all citations to a given mouse-article involve the use of that specialized research mouse in a follow-on experiment, and that most researchers routinely include a citation to the original mouse-article when a particular mouse is used in a follow-on project. Third, both the Cre-Lox and Oncomouse NIH agreements occurred well after the publication dates of a large number of Cre-Lox and Onco mouse-articles; as such, for each mouse-article, we are able to observe citations both before and after the NIH agreement (and compare this to the pattern observed for our control groups which were unaffected by the NIH agreements). Finally, while there was pressure on the NIH and DuPont to move towards a more open regime, both the timing and extent of the openness shock are arguably 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 – are likely to have been drawn from a sample of similar scientific quality/importance but they only differ insofar as the patent over knock-out technology was unenforced by the University of Utah.

Taken together, this empirical approach allows us to exploit the timing of the openness shocks to observe pre- and post-shock citation rates to the treated mouse-articles (those associated with Cre-lox and Onco mice). By also including untreated mouse-articles (those associated with Knock-out and Spontaneous mice), we can more precisely identify a counterfactual estimate of the citation rate that would have occurred if the NIH agreement has not been signed. By measuring citations to Cre-lox and Onco mouse-articles before and after the openness shocks (and by measuring the citations to mouse-paper articles unaffected by the MoUs) we can separately identify the causal impact of both the Cre-lox and Onco openness agreements.

Our estimation approach uses an annual count of forward citations to a given mouse-article. As a starting point, we use a negative binomial to accommodate the fact that citation data comes in the form of skewed count data. Given the heterogeneity among scientific research articles, the nonlinear evolution of citations over time elapsed since initial publication, and the potential for differences over time in citation practices, we include article, article-age and calendar year fixed effects. To address the incidental parameters problem, we estimate a conditional fixed effects estimator (Hausman, Hall & Griliches 1984). While the precise functional form will depend upon the precise test, it is useful to illustrate our overall approach by presenting the key estimation equations.

Our dependent variable is $Citations_{jt}$ which measures the number of citations to a given mouse-article in a given calendar year. Our main specifications include an article fixed effect (γ_j, which is conditioned out in estimation), year effects (β_t) and article-age effects ($\delta_{t-PubYear}$). We then include two measures to

capture the impact of the NIH policy: $OverallWindow$ and $PostOverallShock$. $OverallWindow$ is a dummy variable equal to one for those mouse-articles impacted by a shock in the year of -and the year after- the openness shock. [15] $PostOverallShock$ is the key treatment variable and is equal to one for mouse-articles impacted by the shock in citation-years after the window period (i.e after 1999 for Cre-Lox mouse-articles, and after 2000 for Oncomouse mouse-articles). Using a dataset composed of citations to mouse-articles impacted by the shock and mouse-articles that are unaffected by the shock, our main specification for the impact of the NIH agreements on the *level* of citations is thus:

$$\begin{aligned} & Citations_{jt} \\ = \; & f(\varepsilon_{jt}; \gamma_j + \beta_t + \delta_{t-PubYear} + \Psi_0 OverallWindow_{jt} \\ & + \Psi_1 PostOverallShock_{jt}), \end{aligned}$$

This specification tests for the impact of the NIH agreements by estimating how the citation rate for a mouse-article changes after it is impacted by one of the NIH agreements, accounting for fixed differences in the citation rate across articles and relative to the non-parametric trend in citation rates for the non-treated Knock-out and Spontaneous mouse-article control groups. This approach directly accounts for heterogeneity in the underlying quality of individual articles and the evolution of citations over time.[16]

We then build on this baseline specification to estimate the impact of openness on the nature and diversity of follow-on citations. To evaluate the impact of the openness shocks on different types of citations, we estimate the parameters from a two-equation system that divides $Citations$ into two (mutually exclusive) types and estimates the effects of openness on each type. For example, a key prediction of the model is that openness is associated with an increase in the number of different researchers who utilize a given specialized research mouse. To test this hypothesis, we can contrast the impact of a shift in openness on follow-on publications by authors who have (or have not) cited a particular mouse-article before:

[15] Consistent 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.

[16] It is also possible to separate out the treatment effect in several different ways. Our empirical work includes several specifications that estimate the impact of each NIH agreement separately by including separate window and treatment variables for the Cre-Lox shock and Oncomouse shocks. Also, we are able to evaluate the short-term versus long-term impact of the openness shocks by creating a measure a short-term treatment measure that captures the impact on citations during the first three years after the window period, and a separate long-term treatment measure that captures the permanent impact on citations for citation-years more than four years after the initial treatment.

$$
\begin{aligned}
& NewAuthorCitations_{jt} \\
= \; & f(\varepsilon_{jt}; \gamma_j + \alpha_{NEW-OLD} t + \beta_t + \delta^{NEW}_{t-PubYear} \\
& + \Psi^{NEW}_{CRE_0} CreLoxWindow_{jt} + \Psi^{NEW}_{CRE_1} PostCreLoxShock_{jt} \\
& + \Psi^{NEW}_{ONCO_0} OncoWindow_{jt} + \Psi^{NEW}_{ONCO_1} PostOncoShock_{jt},
\end{aligned}
$$

and

$$
\begin{aligned}
& OldAuthorCitations_{jt} \\
= \; & f(\varepsilon_{jt}; \gamma_j + \beta_t + \delta^{OLD}_{t-PubYear} \\
& + \Psi^{OLD}_{CRE_0} CreLoxWindow_{jt} + \Psi^{OLD}_{CRE_1} PostCreLoxShock_{jt} \\
& + \Psi^{OLD}_{ONCO_0} OncoWindow_{jt} + \Psi^{OLD}_{ONCO_1} PostOncoShock_{jt},
\end{aligned}
$$

We impose several parametric restrictions due to data constraints, including setting the mouse-article fixed effects equal across both equations and allowing the calendar year fixed effects to differ only as a linear function of time. Notably, we allow for the publication age fixed effects to vary freely across both equations, as the evolution of citations in the time 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). Our hypothesis test focuses on whether$\Psi^{NEW}_{CRE_1}$ and $\Psi^{NEW}_{ONCO_1}$ are significantly larger than $\Psi^{OLD}_{CRE_1}$ and $\Psi^{OLD}_{ONCO_1}$, respectively. In other words, we evaluate whether the change in citations occurring after the openness shock arises due to a particular increase in citations by authors who had not previously cited a particular mouse-article. We then develop similar specifications for several citation margins that capture the notion of diversity across research lines described in our theory: citations from new versus old institutions, using new versus old key words, and published in new versus old journals. Similarly, we explore the research response to the openness shocks along a given research line by comparing citations in applied versus basic journals.

In all our analyses, we provide in brackets the coefficients as incidence-rate ratios (a coefficient equal to one implies no effect on $Citations_{jt}$, whereas a coefficient equal to 1.50 implies a 50% boost to $Citations_{jt}$). All models also include and report block bootstrapped standard errors, clustered by mouse-article (MacKinnon, 2002).

5 Empirical 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 linked to an original scientific publication (thereby establishing the population of mouse-articles). Of this large population, we focus only on mouse-articles published between1992 and 1998 (the date of the first NIH agreement). As outlined above, we sample all mouse-articles for four types of mouse engineering technologies defined by MGI: Cre-lox, Onco, Knock-out and Spontaneous mouse-articles (whose production relies upon random mutation rather than engineered genetic manipulation). In total, our sample includes 2638 novel mice linked to 2223 unique mouse-articles. The breakdown is as follows: 52 Cre-lox mice, 160 Onco mice, 2171 Knock-Out mice, and 255 Spontaneous mice.

For all 2223 mouse-articles we obtained information on publication year from PubMed. We then used Thomson ISI Web of Science to collect all follow-on (forward) citations in academic journals for the year immediately following publication through to the end of 2006, to give a total of 525,865 citations. Each citation included detailed information on last author, reprint author, institutional addresses, key words, and journal characteristics (including name, impact factor and a basicness score). The citations were aggregated into 27,442 citation-year observations by combining all the citations received by a given mouse-article in any particular year as the basis for our analysis of the impact of the NIH agreement on the *level* of follow-on innovation.

To capture a variety of measures of the *type* of follow-on innovation, we coded each of the citations according to a set of mutually exclusive categorical variables. Following our theoretical predictions, we focus on margins intended to capture horizontal experimentation across new lines. Our measures include the diversity of researchers in follow-on innovation - new researchers and new institutional affiliations, as well as the diversity of research - new key words and new journals. To illustrate the construction of these variables, take the case of new key words. We consider a citation to include one or more new key words if the key word has never been used in citations to a particular mouse-article in any prior year. A citation is coded old in all other instances. This construction allows us to capture changes in the research landscape. Overall, we generate four new/old categorical variables:

i. New/Old Last Author: defined as new if the last author (listed in ISI Web of Science) has never appeared as a last author before in a citations 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 following a schema developed by Lim (2000). This allows us to capture the predictions of our model regarding the impact of openness

on the vertical change in follow-on innovation i.e. whether these shifts lead to research further along particular research lines (towards commercialization).[17] It is worth noting that in this analysis, multidisciplinary journals are classified as "basic" thus adding a conservative bias against finding an increase in basic research compared to applied research.

Taken together these measures allow us to explore the detailed predictions of our theory regarding the ways in which changes in openness impact the type of follow-on research along both horizontal and vertical dimensions. We implement this analysis following the econometric approach laid out above. Specifically, rather than using all citations in a given year to create the citation-year dependent variable, we group the citations in each citation-year for any mouse-article into two mutually exclusive citation-year observations e.g. new key word citations and old key word citations, basic journal citations and applied journal citations etc. This gives us 54,884 citation-year observations and in each case allows us to examine the impact of the NIH agreements on the two distinctive margins. By comparing the causal impact of the NIH agreements on these two margins, we can investigate the hypothesis that changes in openness create more diverse lines of research, pursued by a more diverse range of scientists. We also investigate where along the research line (from basic to applied) the additional research is taking place. One caveat is worth noting. We do not examine the impact of openness on the public/private citation margin. First, the openness shocks in our analysis are directed specifically to public-sector researchers. Second, for our entire sample we find that 97.5% of all forward citations have at least one of their authors in public institutions (of which 92.5% are only public and only 5% are public-private mix). With only 2.5% of citations having all private-sector authors, this margin is insignificant in the field of engineered mice.

5.2 Variables and summary statistics

Our empirical analysis focuses on measuring follow-on innovation as proxied by citations to the 2223 mouse-articles. Table 1 provides variable names and definitions and Table 2 reports summary statistics for our data. Our mouse-articles are published between 1992 and 1998 (mean = 1995) and have an average of seven authors each. We trace citations to each mouse-article from the year after its publication until 2006 (with the mean of $CitationYear_{jt}$ being 2001), giving us 27,442 citation-year observations. The papers receive a mean of 231 $TotalCitations_j$ between the year following their publication and 2006. Our key dependent variable in the initial set of regressions is $AnnualCitations_{jt}$ measuring the total number of citations to article j in year t. The average number of citations for our mouse-articles is 18.32 (with a minimum of 0 and

[17] Our Basic/Applied Journal definition is based on work by Lim (2000) 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 (see Hicks 1996, for more details)" (Lim 2000 p. 129).

maximum of 336 citations received in any year). This is higher than the mean in other samples of life science papers (e.g. Murray & Stern 2007), highlighting the importance of mouse genetics research in this period.

TABLE 1 HERE
TABLE 2 HERE

In our core analyses we break up the annual citation count for any mouse-article into categorical margins of interest. As outlined above, to measure the diversity of citing authors, we construct the following two dependent variables: $NewAuthorCitations_{jt}$ and $OldAuthorCitations_{jt}$ by measuring number of citations by new (last) authors to article j in year t; and the number of citations by old (last) authors to article j in year t, respectively (mean = 11.0 and 3.7 respectively). We then create an additional new/old dependent variable: $NewInstitutionCitations_{jt}$ and $OldInstitutionCitations_{jt}$ (mean = 16.6 and 9.7 respectively) to capture diversity at the institutional-level. Likewise, to capture diversity across research lines we code citations with new and old key words as $NewKeywordCitations_{jt}$ and $OldKeywordCitations_{jt}$ (mean = 32.8 and 11.0 respectively) as well as citations in new and old journals as $NewJournalCitations_{jt}$ and $OldJournalCitations_{jt}$ (mean = 7.5 and 5.9 respectively). Following similar logic, we capture vertical shifts in research along particular research lines through the dependent variables $BasicCitations_{jt}$ and $AppliedCitations_{jt}$, which measure the number of citations in basic journals to article j in year t; and the number of forward citations in applied journals to article j in year t, respectively (mean = 8.725 and 6.947 respectively).

As described in our empirical specification, we create three shock variables. The first is the $PostOverallShock_{jt}$, equal to one if the article j is subject to either of the two MoU openness shocks, and if the citation year is after the window period for the shock (mean = 0.0482). The second and third variables capture the specific Cre and Onco shocks: $PostCreLoxShock_{jt}$ is equal to one if the article j is subject to the Cre-lox MoU openness shock, and if the citation year is after the Cre-lox window period for the shock (mean = 0.013) and $PostOncoShock_{jt}$ equal to one if the article j is subject to the Onco MoU openness shock, and if the citation year is after the $OncoWindow$ period (2001 or later) (mean = 0.035).

TABLE 3 HERE

6 Results

Our empirical analysis estimates the causal impact of the openness shocks exemplified by the Memorandum of Understanding signed by DuPont, NIH and JAX dramatically opening up the access to Cre-lox (1998) and Onco (1999) mice for academic researchers. Recall that these agreements both reduced downstream expropriation of follow-on innovators (in the case of Cre-lox and Onco) by decreasing the reach-through rights available to DuPont, and increased access for

follow-on innovators to the mice themselves (particularly in the case of Cre-lox mice). Our approach is to observe the annual citations to mouse-articles linked to Cre-lox and Onco mice in the pre- and post- shock period. By comparing the citation patterns to Knock-Out mice and Spontaneous mice unaffected by the MoUs and to the pre-shock trends for the treated mice, we can identify the impact of the shocks to openness.

Our analysis proceeds in several stages. First, we investigate the impact across both shocks on the overall flow of citations received by our mouse-articles. We also decompose the shocks to determine the specific impact of the Cre-lox and Onco shocks to better characterize their different causal impact. In both cases we also examine the time dynamics of the shocks. We then turn to the core of our analysis which first examines the Overall, Cre-lox and Onco shocks on the horizontal flow of research – by different researchers and across research lines and then the vertical flow of research along a given line (from basic to applied) . We capture the horizontal margin of "new" compared to "old" categories of citations, specifically key words, journals, authors, and institutions. In contrast, we use the vertical margin of basic versus applied journals to capture the downstream nature of research. By analyzing the impact of openness within the differences-in-differences framework, we are particularly interested in coefficient on the "shock" variable as this captures the change in citations (overall or for a particular margin) in the pre- and post-shock period. We focus on incidence-rate ratios in our presentation because they are easily interpreted: an incidence-rate ratio, or IRR, provides the multiplicative effect on the expected number of citations received with a one unit change in a regressor (i.e., the null hypothesis of no effect yields a coefficient of 1.0). For example an IRR of 1.25 on the shock variable can be interpreted as a 25% boost in citations in the post shock period.

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

Our regression results begin in Table 4 with a negative binomial specification using $TotalCitations$ as the dependent variable. All specifications use the full set of fixed effects. Equation (4-1) column represents our baseline model, with the $PostOverallShock$ variable. After accounting for the window period, we find that the coefficient on $PostOverallShock$ is significant. On average, mouse-articles affected by the shocks (Cre-lox and Onco mouse-articles) received an additional 21% increase in their annual citation rates after the MoUs are signed. The effect is identified both from the large set of control mouse-article papers and from the pre- and post- variations in article ages. Under specification (4-2) we divide the primary explanatory variable into $PostOverallShock_{S}hortRun$ and $PostOverallShock_{L}ongRun$ but make no other changes to the analysis. We find that the boost in overall citation rates is significant in both periods and is actually growing over time, with a 15% increase through 2003 and a 32% increase for 2004-2006. More than simply a lag in publishing after the initial $PostOverallShock$ period (which is accounted for with the window variable), the significant and increasing boost in both periods represents a positive

feedback effect, where the initial boost focuses greater attention on the lines of research affected by the shocks, resulting in even higher citation rates in the next round of scientific articles.

In (4-3) we repeat these analysis but make separate estimates for the coefficients on the Cre-lox and Onco shocks - a specification that more accurately captures the differences in the two shocks (with respect to openness in the pre period). In (4-3), we show that the *PostCreLoxShock* variable is associated with a statistically significant (but noisy) increase of 18% in citations for Cre-lox mouse-articles compared to 21% for the *PostOncoShock* variable. These results can also be more clearly observed in Figure 1 which provides a graphical representation on the coefficients of the citations to the Onco, Cre-lox and Control (Knock-out and Spontaneous) mouse-articles by year. As the graph highlights, the control mice show a slow, steady upward trend in their annual citations, suggestive of the fact that over time the entire area of mouse genetics is becoming more salient and the source of a wide range of new research lines. However, the coefficients on the Cre-lox and Onco mice are calculated having accounted for this trend and we find that the Cre-lox mice show a similar slow trend up until 2001, after which they take off quite dramatically. This relatively long lag is consistent with the idea that open access to such mice initiated a substantial new research but that it took time for researchers to respond and change their research direction. nonetheless the sharp upward trend is consistent with the notion that the Cre shock was a very sharp and clean natural experiment. In contrast, the Onco mice citation trend is much noisier, including some pre-shock anticipation and then a lag (similar to cre mice) until 2002, at which point they also experience a surge in citations.

These results provide strong support for one key claim of this paper – that positive shocks to openness foster research intensity, rather than hindering it because of appropriability concerns surround critical research outputs. This adds support to previous empirical results, for example by Furman and Stern (2006), showing that the deposit of individual cell-lines (which provides openness through formal access) also increases follow-on innovation. In a complementary result, Murray and Stern (2007) find that limits on openness with the grant of intellectual property rights over knowledge have the converse effect; it decreases follow-on citations. Taken together, these results highlight the sensitivity of follow-on researchers to a variety of openness conditions, and provides increasing support for the perspective that these results are driven by researchers shifting their research choices rather than shifting their citations – it is hard to imagine the research community being so strategic in their citations that they increase and decrease their citations according to the precise timing and degree of openness shocks. Furthermore, our results on temporal dynamics are consistent with our theoretical setup, specifically the multi-staged view of innovation: if openness leads to more research activity and potentially to a branching out of new research lines (a conjecture we test in our next set of regressions) then these new lines would themselves generate follow-on research activity, amplifying over time the effects of any shocks to openness.

TABLE 4 HERE
FIGURE 1 HERE

To examine the impact of the openness shocks on the horizontal expansion of follow-on research, and to move to specifications that capture our core theoretical insight - that openness will have a more significant impact on new, early-stage research lines, where openness is complementary with freedom - we examine the impact of the openness shocks on several citation margins. As explained in the Estimated Equations section, we consider a series of two-equation systems that allows us to contrast various margins of the annual citations, helping to clarify the overall changes in behavior.

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

In Tables 5, 6 and 7, we present our analysis for the second main theoretical claim in our model predicting that greater openness will lead to greater horizontal experimentation, spawn a diverse array of new research lines and encourage the participation of new researchers who have previously not contributed to this arena of knowledge. We first present our evaluation of the impact of openness shocks on the diversity of researchers participating in follow-on research. Our key comparison is between researchers listed as the last author (the senior scientist) on citations who have never previously been listed on a citation to the mouse-article of interest, captured in our measure, $NewAuthorCitations$, and those researchers who have been previously listed in a citation to the particular mouse-article: $OldAuthorCitations$. In the stacked regressions presented in (5-1a) and (5-1b) we estimate whether the marginal impact of the $PostOverallShock$ is different for new versus old last-authors. When we separately evaluate the Cre-lox and Onco shocks on new and existing authors (5-2a) and (5-2b), we find that the Cre-lox openness shock leads to a 25% increase in new last-author citations, with no increase in old last-author citations. Similarly, the Onco shock leads to a 22% increase in new-author citations. Turning to the time dynamics for the overall shock (5-2a and 5-2b), we find an 18.5% increase in citations by new authors, compared to statistically insignificant increase in citations by old authors (with the difference of the coefficients significant at the 1% level), and 36% versus 21% for new versus old authors in the long-run. This provides strong evidence for the hypothesis that an increase in openness leads to new lines of research, as the shocks led to new authors focusing on the field.

In the final set of specifications in Table 5 (5-4a and 5-4b), we turn to an alternative measure of the diversity captured by the institutional affiliation. In this case institutions are coded from the address field of the particular mouse-article citation. This is particularly informative because it allows us to explore the micro-foundations of openness and mouse exchange at the institutional level. If researchers within a given institution (e.g. Northwestern University) share mice freely with one another once one of their colleagues has made the investment in accessing a mouse (or engineering one) then we

would expect the surge in new authors to come predominantly from new institutions. Furthermore, any university-level agreement made prior to the MoU made follow-on research possible for all scientists within the university. As throughout Table 5, we used stacked regressions to estimate specifications comparing $NewInstitutionCitations$ and $OldInstitutionCitations$. Comparing (5-4a) and (5-4b), the impact of the overall openness shock increases citations from new institutions by 20% compared to 14% from old (existing) institutions. In other words, while the effect is less dramatic than the increased diversity of authors, the boost in marginal citations does accrue (significantly) to authors affiliated with new institutions.

TABLE 5 HERE

While our theoretical predictions highlight the importance of openness on reducing the fixed cost of critical upstream inputs into research projects, another important aspect of openness is the degree to which it facilitates horizontal experimentation by researchers now free to match with a variety of ideas, particularly given the conditions of freedom existing in the academic sector that we examine here. We capture this horizontal diversity using the measure of key words represented in a particular citation. Recall that these key words are defined by the cataloguing service (ISI Web of Science) and therefore not subject of strategic intervention by researchers themselves. We compare the citation margin between $NewKeywordCitations$ and $OldKeywordCitations$ in (6-1a) and (6-1b) finding that the $PostOverallShock$ is 25% for new key words and and insignificant for old key words. This confirms our prediction that openness does indeed have a substantial impact on the diversity of new research lines. When we include the time dynamics (6-2a) and (6-2b) we find that the short run $PostOverallShock$ effect on new key words is 20%, and increases to 35% in the long run (both are significant at the 1% level). The old key word impact is insignificant. Taken together these provide strong evidence for expanding research lines. When we decompose the openness shock into the Crelox and Onco shocks, the results are also dramatic. The $PostCreLoxShock$ is 30% while the Onco shock is only 20% (significant at the 5% and 10% level respectively) suggesting that it is the Cre-lox shock that has the most salient impact on the initiation of diverse early-stage lines. Neither the Cre-lox nor the Onco Shocks have a significant impact on old key words.

TABLE 6 HERE

Our final investigation to establish the impact of openness on diversity is the emergence of research lines focusing on new areas of scientific study captured in journals. As a proxy for this breadth of research, we compare the citation margin between $NewJournalCitations$ and $OldJournalCitations$, where a "new" journal is one which has never before published an article citing the original mouse-paper article in question. In Table 7, we see that the $PostOverallShock$ in (7-1a) and (7-1b) leads to a 24% increase (significant at the 1% level) in citations from new journals, and no significant increase in citations from old

journals. We further investigate the impact of openness in (7-2a) and (7-2b) which show that the short run effect is 22% for new journals increasing to 27% in the long run, while there is no short run impact for old journals, but the long run *PostOverallShock* is 23%. Finally, in (7-3a) and (7-3b) we examine the Cre and Onco shocks, finding that the Cre-lox shock has an impact on *NewJournalCitations* of 24% but the effect is noisy and only significant at the 15% level, however, the Onco shock is leads to a 24% increase to citations in new journals with no significant increase in citations in old journals (significant at the 1% level).

TABLE 7 HERE

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

We now turn to the effects of openness shocks on the vertical distribution of research, in other words, whether openness shocks move research along any particular line towards later stage projects. We do this by examining the marginal impact of the openness shocks on the production of research in basic versus applied research journals. Recall that these categories are determined by examining the journal in which citations are published, categorized according to how close to clinical application the work typically published (across the entire stock of articles published in the journal over a five year time period). In (8-1a) and (8-1b), we find that the *BasicCitations* dependent variable increases 23% during the post-shock period; at the same time, the *AppliedCitations* variable experiences 18% increase during the post-shock period. This suggests that across both shocks, the average impact accrues to both basic and applied citations. In our next regressions, however, we provide deeper insights into these patterns by again considering the contrasting natures of the Cre-lox and Onco shocks and disentangling their distinctive implications. Recall that in the pre-shock period, not only were there stringent reach-through rights associated with Cre-lox mice, but also very limited access as ex ante enforcement of IP rights had limited their circulation and exchange. In contrast, Onco mice were made available through JAX - although these researchers remained concerned that if they found interesting commercial applications they may be subject to ex post IP enforcement. As a result, the Onco shock also reduced reach-through rights but had a more limited impact on access. The specifications in (8-2a) and (8-2b) reveal that the Cre-lox shock is concentrated in basic citations, while the Onco shock has a significant effect only on applied citations. Specifically, the Cre-lox shock leads to a dramatic 78% increase in basic citations during the post-shock period, but has a 21% decrease on the applied-research citation flows (significant at the 10% level). By contrast, the impact of the Onco shock is concentrated in the more applied research stages and leads to a 56% increase during the period through 2006 for applied citations; at the same time, the Onco shock has no significant impact on basic citations. This is consistent with the view that when upstream access is already secured (as in the case for Onco mice), then an agreement that

shifts the balance of appropriability toward follow-on innovators and away from the initial innovator (DuPont), then there is more applied research.

TABLE 8 HERE

These results are further reinforced when we look at the time dynamics (8-3a) and (8-3b). In this case, rather than look at the time dynamics for the overall shock, we examine the time dynamics for the Cre-lox and Onco shocks separately. We find that the Cre-lox shock has a 63% increase in basic citations in the short run and a dramatic 114% increase over the next three years (through 2006). There is a significant negative impact on applied citations in either the short run (25%) but no change in the long run. While we might have anticipated that there would be a gradual shift to applied research in the long run, this suggests that the early stages of the Cre-lox research lines take time and that applications are relatively far away. Conversely, in the Onco case, the shock to citations is entirely concentrated in applied research with a 51% boost in the short run and 63% in the long run.

Taken together, our findings suggest that both the Cre-lox and Onco shocks had an important impact on the rate and nature of follow-on innovation. Of course our interpretation depends upon the extent to which the MoU shocks to openness were truly exogenous. After all, they reflected the endogenous choice of DuPont, JAX, and the NIH. There is, however, strong evidence to suggest that the Cre-lox shock and (to a lesser extent) the Onco shock were unanticipated in their timing and terms by the scientific community and that while the academic community had agitated for broader access, this had been a continuous request starting in the early 1990s, rather than a significant sea change in response to changing technical opportunities (Murray, 2008). Moreover, our focus is on the behavioral (citation) response of over 5,000 follow-on researchers who were not part of the intense, but largely private, negotiations. More than simply a policy announcement, or even an agreement that ratified behavior already taking place, the MoUs directly and dramatically changed the openness of a set of key research inputs.

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 regarding certain genetically engineered mice.

Overall we found, perhaps not surprisingly, that there is an increased overall level of follow-on research taking place after the NIH-DuPont-JAX openness agreements. Building on this initial result, we explored the particular margins

where this increased innovation is taking place, developing measures of innovation that allowed us to test the predictions of our theory. First, we obtained robust evidence that increased openness was associated with the exploration of a wider range of more diverse research paths i.e. horizontal experimentation. This finding highlights a feature of early stage knowledge overlooked in many of the current models of innovation - the fact that it is non-rivalrous and as a consequence, can, in principle, be applied across multiple later-stage domains and applications. Second, when we compared the impact of openness on horizontal exploration versus vertical exploitation we found that on balance, when pre-existing IP restrictions limited access to research materials (rather than simply serve as a threat of potential future enforcement), the main impact of openness is concentrated in an increase in more basic and more high-quality follow-on research publications. In contrast, when prior arrangements (informal or formal) have allowed for access even with some threat of enforcement, the openness shock is concentrated in more applied follow-on research.

Our results highlight that the current literature on intellectual property and innovation has neglected a key potential cost of intellectual property - the limits that IP rights may place on the diversity research that would otherwise be pursued by follow-on innovators taking a single powerful idea and experimenting across multiple research lines.

Our results also have strong implications for the organization of research and its contribution to innovation and growth both in academia and the private sector. For nations such as China, for example, who seek to increase knowledge production through greater funding and an emphasis on incentives to publish in academia, we argue that without a commitment to openness as well as freedom, these investments are unlikely to be effective (see Murray 2007 for a related discussion). This commitment to openness requires careful consideration and must be balanced against the current enthusiasm for IP rights in academia. The prevailing view on openness (and IP rights) is shaped by the technology transfer model of the United States as structured by the 1980 Bayh-Dole Act. By placing IP rights in the hands of the universities Bayh-Dole allowed them to shape the ways in which their IP was enforced upon follow-on innovators. The goal was to provide key incentives for follow-on exploitation and the transformation of basic research investments into commercial products. Our results highlight one of the possible dangers of excessive IP enforcement: if IP is used to restrict openness particularly at very early stages of the research line, then it is possible that the rich array of exploration projects that are key to diverse follow-on innovation will be stifled. In practical terms, there are a number of ways of managing IP and access rights to try and maximize horizontal exploration and vertical exploitation. However, this will require policy makers, university administrators and academics themselves to pay greater attention to the organization of research, particularly the terms and conditions that pertain to access to patented research inputs, but also more broadly, the institutions that enhance openness.

Lastly, these results should affect the way we think about the role and importance of IP protection throughout the innovation process in the private sector. In

particular, our framework suggests that more attention be paid by economists to recent attempts by the corporate sector to generate new sources of profit built on the openness of knowledge production by others (Huang & Murray 2008). Thus, Tapscott and Williams (2006) explain how IBM has managed to recover from competition with Microsoft by engaging in the openness promoted by Linux. More generally, a systematic analysis of the forces and trade-offs at work in an economic environment with both proprietary and open firms competing with each other, awaits future research.

References

[1] Aghion, P, Dewatripont, M, and Stein, J (2008), "Academic Freedom and Private Sector Focus", mimeo, Harvard, forthcoming, *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] Anderson, Christopher. 1993. "Researchers win decision on knockout mouse pricing". Science, Vol. 260 (5104), pp. 23-24.

[4] Azoulay, Pierre, Waverly Ding, and Toby Stuart. 2007. "The Impact of Academic Patenting on (Public) Research Output. Journal of Economic Behavior and Organization 63, no. 4, pp. 599-623.

[5] Boguski, Mark S. 2002. "Comparative genomics: the mouse that roared." Nature 420(6915): 515-6.

[6] Bresnahan, Timothy F. & Trajtenberg, M., 1995. "General purpose technologies 'Engines of growth'?," Journal of Econometrics, Vol. 65(1), pages 83-108

[7] Brinster, Ralph L., H. Y. Chen, M. Trumbauer, A. W. Senear, and R. Warren et al. 1981. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. Cell 27:223-231.

[8] Brinster, Ralph L., H.Y. Chen, A. Messing, T. Vandyke, A.J. Levine and R.D. Palmiter. 1984. "Transgenic Mice Harboring SV40 T-Antigen Genes Develop Characteristic Brain-tumors. Cell 37 (2): 367-379.

[9] Constantini, Frank and E. Lacy, 1981. "Introduction of a rabbit beta-globin gene into the mouse germ line". Nature 294:92-94.

[10] Dasgupta, Partha and David, Paul. 1994. "Towards a New Economics of Science" Research Policy (23), 487-521.

[11] David, Paul. 2001a. "From Keeping 'Nature's Secrets' to the Institutionalization of 'Open Science" Stanford Working Paper #01-006.

[12] David, Paul. 2001b. "Will Building Good Fences really make Good Neighbors in Science" Stanford Working Paper #01-005.

[13] de Solla Price, D. J. 1965. Networks of scientific papers. Science. 149(3683) 510-515.

[14] Ding, Waverly, Fiona Murray, and Toby Stuart. 2006. "Gender Differences in Patenting in the Academic Life Sciences." Science, Vol. 313 (5787), p.665-667.

[15] Dreyfuss, Rochelle. 2004. "Protecting the Public Domain of Science: Has the Time for an Experimental Use Defense Arrived?" 46 Arizona Law Review p.457.

[16] Fauchart, E. E. Von Hippel. 2007. Norms-Based Intellectual Property Systems: The Case of French Chefs. Organization Science.

[17] Furman, J., S. Stern. 2006. Climbing atop the shoulders of giants: The impact of institutions on cumulative research. Working paper, National Bureau of Economic Research, Cambridge, MA.

[18] Garfield, E. 1979. Citation Indexing. Philadelphia, PA: ISI.

[19] Gordon, J. W., G. A. Scangos, D. J. Plotkin, J. A. Barbosa, and F. H. Ruddle, 1980. "Genetic transformation of mouse embryos by microinjection of purified DNA". Proc. Natl. Acad. Sci. USA 77:7380-7384.

[20] Hagström, Warren O. 1965. The scientific community. New York: Basic Books.

[21] Hall, B. H., A. B. Jaffe, M. Trajtenberg,. 2001. The NBER patent citation data file: Lessons, insights and methodological tools. Working paper no. 8498, NBER, Cambridge, MA.

[22] Heller, Michael. 2008. Gridlock Economy: How too much ownership wrecks markets, stops innovation and costs lives. Philadelphia PA: Basic Books.

[23] Heller, M, and Eisenberg, R (1998), "Can Patents Deter Innovation? The Anticommons in Biomedical Research", *Science,* 280, 698-701.

[24] 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.

[25] Howitt, P (2003), "The Economics of Science and the Future of Universities", The Timlin Lecture 2000, University of Saskatchewan.

[26] Huang, Kenneth, F. Murray. 2008. Does Patent Strategy Shape The Long-Run Supply Of Public Knowledge? Evidence From The Human Genome. Forthcoming in Academy of Management Journal.

[27] Marshall, Eliot. 1999. "New NIH Rules Promote Greater Sharing of Tools and Materials". Science, Vol. 286(5449), pp. 2430-2431.

[28] Marshall, Eliot. 2000. "NIH Cuts Deal on Use of Oncomouse". Science, Vol. 287(5453), p. 567.

[29] Marshall, Eliot. 2002. "DuPont Ups Ante on Use of Harvard's Oncomouse". Science, Vol. 296, pp. 1212-1213.

[30] Merton, Robert. 1973. The Sociology of Science: Theoretical and Empirical Investigations. Edited by Norman Storer. Chicago: University of Chicago Press.

[31] Mokyr, Joel. 2002. The Gifts of Athena: Historical Origins of the Knowledge Economy. Princeton: Princeton University Press.

[32] Mowery, David, Richard Nelson, Bhaven Sampat, and Arvids Ziedonis. 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.

[33] Murray, Fiona and Scott Stern. 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, Vol. 63:4, pp. 648-687.

[34] Murray, Fiona. 2002. "Innovation as co-evolution of scientific and technological networks: exploring tissue engineering". Research Policy, Vol. 31, Issues 8-9, pp: 1389-1403.

[35] Murray, Fiona. 2007. "The Stem Cell Market: Patents & the Pursuit of Scientific Progress." New England Journal of Medicine, Vol. 356:23, pp. 2341-2343.

[36] Murray, Fiona. 2008. The Oncomouse that Roared: Hybrid Exchange Strategies As A Source Of Productive Tension At The Boundary Of Overlapping Institutions. Forthcoming American Journal of Sociology and Working Paper, Sloan School of Management.

[37] Nelson, Richard R. 1959. "The Simple Economics of Basic Scientific Research," Journal of Political Economy 67(3), 297-306.

[38] O'Mahony, S. 2003. "Guarding the Commons: How Community Managed Projects Protect Their Work", Research Policy: 32, 1179-1198.

[39] Oliar, D., C. Sprigman. 2008. Intellectual Property Norms Among Stand-Up Comedians, Va. L. Rev.

[40] Owen-Smith, Jason, and Walter W. Powell. 2003. "The Expanding Role of University Patenting in the Life Sciences: Assessing the Importance of Experience and Connectivity." Research Policy 32(9): 1695-1711.

[41] Paigen, Ken. 2003. One Hundred Years of Mouse Genetics: An Intellectual History. I. The Classical Period (1902–1980). Genetics 2003 163: 1-7.

[42] Rader, Karen. 2004. Making Mice. Princeton, NJ: Princeton University Press.

[43] Russell, E. S. (1978). "Origins and history of mouse inbred strains: contributions of Clarence Cook Little". In Origins of Inbred Mice, Morse, H. C., eds. (Academic Press, NY), pp. 33-43.

[44] Rysman, M., & Simcoe, T. S. 2008. Patents and the performance of voluntary standard setting organizations. Forthcoming Management Science and Working Paper No. 05-22, NET Institute. Available at SSRN: http://ssrn.com/abstract=851245.

[45] Scotchmer, S. 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law", Journal of Economic Perspectives 5, 29-41.

[46] Scotchmer, Suzanne. 1996. "Protecting Early Innovators: Should Second-Generation Products be Patentable?" The Rand Journal of Economics 27, Summer 1996, 322-331.

[47] Stern, S (2004), "Do Scientists Pay to Be Scientists", Management Science.

[48] Stewart, Timothy A., Paul K. Pattengale, and Philip Leder. 1984. "Spontaneous Mammary Adenocarcinomas in Transgenic Mice That Carry and Express MTV/myc Fusion Genes". Cell 38: 627-637.

[49] Strong, Lionel C. 1978. Inbred mice in science. In Origins of Inbred Mice, Morse, H. C., eds. (Academic Press, New York), pp. 45-67.

[50] Wagner, E. F., Stewart, Timothy A., and Beatrice Mintz. 1981. "The human beta-globin gene and a functional viral thymidine kinase gene in developing mice." Proc Natl Acad Sci 78, 5016-20.

[51] Wagner, T.E., P. C. Hoppe, J.D. Jollick, D.R. Scholl, R.L. Hokinka et al., 1981 Microinjection of a rabbit beta-globin gene into zygotes and its subsequent expression in adult mice and their offspring. Proc. Natl. Acad. Sci. 78:6376-6380.

[52] Walsh, John, Ashish Arora, and Wesley Cohen. 2003. "Science and the Law: Working Through the Patent Problem." Science, Vol 299, Issue 5609, 1021 , 14 February 2003

[53] Walsh, John, Charlene Cho, and Wesley Cohen. 2005. "View from the Bench: Patents and Material Transfers." Science, Vol. 309, Issue 5743, 23 September 2005.

TABLE 1: VARIABLES & DEFINITIONS

VARIABLE	DEFINITION	SOURCE
PUBLICATION CHARACTERISTICS		
Publication Yearj	Year in which article *j* is published	PM
# Authorsj	Count of the number of authors of Article *j*	PM
Total Citationsj	# of FORWARD CITATIONS from publication date through 2006	SCI
CITATION-YEAR CHARACTERISTICS		
Annual Citations	# of Forward Citations to Article *j* in Year *t*	SCI
Citation Yearjt	Year in which FORWARD CITATIONS are received	SCI
CITATION CHARACTERISTICS		
New Author Citation	Dummy variable equal to 1 if the last author has not appeared in the citations to the mouse-article in prior years; 0 otherwise	PM
Old Author Citation	Dummy variable equal to 1 if the last author has appeared in the citations to the mouse-article in prior years; 0 otherwise	PM
New Institution Citation	Dummy variable equal to 1 if the institutional affiliation has not appeared in the citations to the mouse-article in prior years; 0 otherwise	PM
Old Institution Citation	Dummy variable equal to 1 if the institutional affiliation has not appeared in the citations to the mouse-article in prior years; 0 otherwise	PM
New Key Word Citation	Dummy variable equal to 1 if the key word has not appeared in the citations to the mouse-article in prior years; 0 otherwise	PM
Old Key Word Citation	Dummy variable equal to 1 if the key word has appeared in the citations to the mouse-article in prior years; 0 otherwise	PM
New Journal Citation	Dummy variable equal to 1 if the publishing journal has not appeared in the citations to the mouse-article in prior years; 0 otherwise	PM
Old Journal Citation	Dummy variable equal to 1 if the publishing journal has appeared in the citations to the mouse-article in prior years; 0 otherwise	PM
Basic Citation	Dummy variable equal to 1 if the publishing journal is identified as a basic-research journal (SOURCE: CHIBasic variable); 0 otherwise	PM
Applied Citation	Dummy variable equal to 1 if the publishing journal is identified as an applied-research journal (SOURCE: CHIBasic variable); 0 otherwise	PM
At Least One Public Author	Dummy variable equal to 1 if *at least* one institutional affiliation associated with the citing article is a university or government organization; 0 otherwise	PM
Private Author	Dummy variable equal to 1 if all institutional affiliations associated with the citing article is a biotechnology or pharmaceutical firm; 0 otherwise	PM
OPENNESS SHOCK CHARACTERISTICS		
Post Overall Shockjt	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
Post Overall Windowjt	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
Post Crelox Shockjt	Dummy variable equal to 1 if Article *j* is associated with the Cre-Lox openness MoU and that agreement is in effect in year *t*.	MGI
Post Crelox Windowjt	Dummy variable equal to 1 if Article *j* is associated with the Cre-Lox openness MoU and that agreement is in its initial period in year *t*.	MGI
Post Onco Shockjt	Dummy variable equal to 1 if Article *j* is associated with the Onco openness MoU and that agreement is in effect in year *t*.	MGI
Post Onco Windowjt	Dummy variable equal to 1 if Article *j* is associated with the Onco openness MoU and that agreement is in effect in year *t*.	MGI

TABLE 2: MEANS & STANDARD DEVIATIONS

VARIABLE	N	MEAN	STD. DEV.	MIN	MAX
PUBLICATION CHARACTERISTICS (N = 2,223 original publication)					
Publication Yearj	2223	1995.35	2.83	1983	1998
# Authorsj	2223	7.034188	3.41921	1	34
Total Citationsj	2223	209.60	231.22	1	2543
CITATION-YEAR CHARACTERISTICS (N = 27,442 citation-year observations)					
Citation Year$_{jt}$	27442	2001.100	3.331	1993	2006
Annual Citations $_{jt}$	27442	18.317	21.132	0	336
New Author Citations $_{jt}$	27442	11.027	13.000	0	243
Old Author Citations $_{jt}$	27442	3.712	5.212	0	58
New Institution Citations$_{jt}$	27442	16.616	17.427	0	287
Old Institution Citations $_{jt}$	27442	9.671	13.346	0	135
New Key Word Citations$_{jt}$	27442	32.782	34.308	0	492
Old Key Word Citations $_{jt}$	27442	11.008	16.235	0	202
New Journal Citations $_{jt}$	27442	70.879	65.864	0	794
Old Journals Citations $_{jt}$	27442	52.252	59.326	0	620
Basic Citation $_{jt}$	27442	8.725	10.942	0	151
Applied Citation $_{jt}$	27442	6.947	10.378	0	157
All Public Authors Citation $_{jt}$	27442	15.115	17.110	0	253
At Least One Private Author Citation$_{jt}$	27442	1.349	2.697	0	45
OPENNESS SHOCK CHARACTERISTICS (N = 27,442 citation-year observations)					
Post Overall Shock$_{jt}$	27442	0.0482	0.2143	0	1
Overall Window$_{jt}$	27442	0.0147	0.1204	0	1
Post Crelox Shock$_{jt}$	27442	0.0133	0.1144	0	1
Crelox Window$_{jt}$	27442	0.0031	0.0552	0	1
Post Onco Shock$_{jt}$	27442	0.0350	0.1837	0	1
Onco Window$_{jt}$	27442	0.0117	0.1074	0	1

TABLE 3: SUMMARY STATISTICS BY MOUSE TECHNOLOGY

		MOUSE TECHNOLOGY			
VARIABLE	**N**	**CRELOX**	**ONCO**	**OTHER GM**	**SPONTANEOUS**
PUBLICATION CHARACTERISTICS (N = 2,223 original publication)					
Publication Year$_j$	2223	1996.549	1991.737	1995.448	1990.789
# Authors$_j$	2223	5.250	5.944	7.341	4.718
Total Citations$_j$	2223	158.831	228.959	234.198	68.411
CITATION-YEAR CHARACTERISTICS (N = 27,442 citation-year observations)					
Annual Citations	27442	15.3340	13.3326	20.9152	3.8202
New Author Citations	27442	10.1294	7.6285	12.5957	2.3799
Old Author Citations	27442	2.6305	2.2984	4.3015	0.6584
New Institutions	27442	15.6910	11.0114	18.9562	3.9763
Old Institutions	27442	8.7286	6.1850	11.1357	1.8031
New Key Words	27442	75.4572	50.7871	80.2499	17.5752
Old Key Words	27442	35.7996	39.9560	59.6379	11.1171
New Journal Citations	27442	7.5511	4.7736	8.5364	1.7752
Old Journal Citations	27442	4.7182	4.6010	6.6681	1.2618
Basic Citations	27442	8.8288	5.0855	9.9965	2.1295
Applied Citations	27442	3.3612	6.4306	7.8953	1.2437
All Public Author Citations	27442	13.2443	10.9772	17.2503	3.1377
At Least One Private Author Citations	27442	0.7724	0.9591	1.5539	0.2583

TABLE 4: IMPACT OF OPENNESS SHOCKS ON ANNUAL CITATION FLOWS

	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 with Overall Shock	**(4-2)** Overall Shock with Time Dynamics	**(4-3)** Baseline Model with Cre & Onco Shocks
Post Overall Shock	**[1.213]***** 0.1934 (0.0507)		
Post Overall Shock Short-run		**[1.152]**** 0.1411 (0.0591)	
Post Overall Shock Long-run		**[1.320]***** 0.2773 (0.0777)	
Post Cre-lox Shock			**[1.178]*** 0.1637 0.0919
Post Onco Shock			**[1.212]***** 0.1921 (0.0610)
Window+ - Overall - Cre - Onco	**[1.119]***** **0.1124** **(0.0405)** **-** **-**	**[1.122]**** **0.1152** **(0.0472)** **-** **-**	 **-** **[0.983]** **-0.017** **(0.123)** **[1.163]***** **0.151** **0.0448**
Parametric Restrictions			
Age FEs = 0			
Year FEs = 0			
Log-likelihood	-67168.977	-67153.037	-67164.516
# of Observations	27428	27428	27428

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

Coefficients for the Window period are included in all regressions but suppressed in order to focus on key variables in the analysis. IRRs reported in brackets; raw coefficients reported in middle line.

TABLE 5: IMPACT OF OPENNESS SHOCKS ON CITATIONS BY NEW VS. OLD 'LAST AUTHORS' & BY 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 With Time Dynamics	**(5-3b)** DV= Old Author Citations With Time Dynamics	**(5-4a)** DV= New Institution Citations	**(5-4b)** DV= Old Institution Citations
Post Overall Shock	[1.250]*** 0.223 (0.054)	[1.082] 0.0785 (0.0789)					[1.202]*** 0.184 (0.0494)	[1.142]** 0.133 (0.0612)
Post Overall Shock Short-run					[1.185]*** 0.170 (0.0538)	[0.994] -0.0056 (0.814)		
Post Overall Shock Long-run					[1.363]*** 0.310 (0.0695)	[1.207]** 0.188 (0.0801)		
Post Cre-lox Shock			[1.251]** 0.224 (0.108)	[0.992] -0.0083 (0.099)				
Post Onco Shock			[1.220]*** 0.199 (0.067)	[1.127] 0.120 (0.0736)				
Parametric Restrictions								
Separate Age FEs = 0								
Common Year FEs = 0								
Log-likelihood								
# of Observations								

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

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 With Time Dynamics	**(6-2b)** DV= Old Key Word Citations With Time Dynamics	**(6-3a)** DV=New Key Word Citations	**(6-3b)** DV=Old Key Word Citations
Post Overall Shock	[1.250]*** 0.223 (0.0738)	[0.977] -0.0230 (0.0732)				
Post Overall Shock Short-run			[1.197]*** 0.180 (0.0586)	[0.926] -0.0765 (0.0666)		
Post Overall Shock Long-run			[1.350]*** 0.300 (0.0843)	[1.052] 0.0504 (0.0784)		
Post Cre-lox Shock					[1.302]** 0.264 (0.104)	[0.894] -0.112 (0.112)
Post Onco Shock					[1.202]* 0.184 (0.0965)	[1.023] 0.0225 (0.115)
Parametric Restrictions						
Separate Age FEs = 0						
Common Year FEs = 0						
Log-likelihood						
# of Observations						

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

TABLE 7: IMPACT OF OPENNESS SHOCKS ON CITATIONS IN NEW VS. OLD JOURNALS

	STACKED NEGATIVE BINOMIAL **[Incident rate ratios reported in square brackets]** **Estimated coefficients in 2nd line.** **(Block bootstrapped SEs reported in parentheses)**					
	(7-1a) DV= New Journal Citations	**(7-1b)** DV= Old Journal Citations	**(7-2a)** DV= New Journal Citations With Time Dynamics	**(7-2b)** DV= Old Journal Citations With Time Dynamics	**(7-3a)** DV= New Journal Citations	**(7-3b)** DV= Old Journal Citations
Post Overall Shock	[1.237]*** 0.213 (0.0711)	[1.108] 0.103 (0.0706)				
Post Overall Shock Short-run			[1.223]*** 0.201 (0.0546)	[1.022] 0.0213 (0.0599)		
Post Overall Shock Long-run			[1.274]*** 0.242 (0.0776)	[1.234]*** 0.210 (0.0764)		
Post Cre-lox Shock					[1.235] 0.211 (0.145)	[1.105] 0.100 (0.133)
Post Onco Shock					[1.236]*** 0.212 (0.065)	[1.108] 0.103 (0.087)
Parametric Restrictions						
Separate Age FEs = 0						
Common Year FEs = 0						
Log-likelihood						
# of Observations						

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

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

	STACKED NEGATIVE BINOMIAL **[Incident rate ratios reported in square brackets]** **Estimated coefficients in 2nd 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	**(8-3a)** DV= Basic Journal Citations with Time Dynamics	**(8-3b)** DV= Applied Journal Citations with Time Dynamics
Post Overall Shock	[1.225]*** 0.203 (0.0732)	[1.184]** 0.169 (0.0766)				
Post Cre-lox Shock			[1.777]*** 0.575 (0.0975)	[0.797]* -0.2269 (0.117)		
Post Onco Shock			[1.029] 0.029 (0.0611)	[1.562]*** 0.446 (0.0739)		
Post Cre-lox Shock Short-run					[1.631]*** 0.4891 (0.0914)	[0.745]** -0.2950 (0.1196)
Post Cre-lox Shock Long-run					[2.140]*** 0.7606 (0.1178)	[0.915] -0.0889 (0.1522)
Post Onco Shock Short-run					[1.030] 0.0298 (0.0756)	[1.514]*** 0.4150 (0.0788)
Post Onco Shock Long-run					[1.029] 0.0290 (0.0861)	[1.632]*** 0.4898 (0.1050)
Parametric Restrictions						
Separate Age FEs = 0						
Common Year FEs = 0						
Log-likelihood						
# of Observations						

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

FIGURE 1: IMPACT OF OPENNESS SHOCKS ON CITATIONS TO ONCO, CRE_LOX & CONTROL MOUSE-ARTICLES