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

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

October 10, 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
[‡]Universite Libre de Bruxelles
[§]Harvard University
[¶]Kellog 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 the incentives for innovation when commercial returns depend upon the allocation of intellectual property rights 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 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 to encourage the establishment of that new research line, 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).

By highlighting a single step-by-step research line, this approach 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 would prefer to pursue (Shane 2001). In other words, rather than focusing exclusively on the value generated along a single line, 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 grid lock arising from an upstream patent "thicket," little attention has been paid to the interaction between the openness of sci-

entific knowledge and the diversity of scientific experimentation. This paper builds on recent research analyzing the distinctive incentives and control rights provided to academic versus industrial researchers (David and Dasgupta 1994, David 2001ab, Stern, 2004), and more specifically on Aghion, Dewatripont & Stein 2007 which emphasizes the role of academic freedom defined as the granting of control rights to researchers. 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 the perception of opportunities or on pure curiosity of individual academic researchers. Here, we use this 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 increase the overall flow of research output, and in particular it is 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 as identified by 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

4

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 role of academic freedom: the ADS framework

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.

5

practical strategy, a researcher may choose to follow an "alternative" strategy in working with an existing idea. 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 $\alpha$. 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 $\alpha$, 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 $\alpha p$.

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.

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 $pV > z$.

Now, let $\Pi_2$ 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 $E(\pi_1^a) = \alpha p\Pi_2 - w_a$.

---

[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.

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 $\Pi_i$ 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.

In ADS, the above result is generalized to the case where the "alternative" strategy consists in working on 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 change the 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 it has a higher probability of generating entirely new research lines than private research does (note that the probability of such an event, possibly the result of an "accidental" discovery, is nonzero for either mode of organization of research).

## 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 two 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 previous ideas and also by making it more likely that the alternative strategies pursued by free researchers will actually lead to new lines. We now formalize 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: restricting hold-up and facilitating downstream transmission

For simplicity, consider a two-stage research line where stages 1 and 2 are managed by the private sector. And suppose that what openness does is to increase the extent to which stage 2 can hold-up and thereby expropriate stage 1. Thus,

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

where $\psi$ measures openness.

The stage-1 value of the line can be written as:

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

Thus, trivially, increasing $\psi$ fosters stage-2 research at the expense of stage 1 research since

$$\frac{d\Pi_2}{d\psi} > 0 > \frac{d\Pi_1}{d\psi}.$$

Now, suppose that openness achieves one more thing, which is to also increase the possibility for stage-1 researcher to transmit her research to stage 2 researcher(s). Thus

$$\Pi_1 = pA(\psi)(\Pi_2 - 2\psi) - w_p = A(\psi)(p^2 V - p\psi) - (1-p)(1+p)w_p,$$

where $A(\psi)$ is an increasing and concave function of $\psi$.

Then we have:

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

and

$$\frac{d^2\Pi_1}{d\psi^2} = A''(\psi)(p^2 V - p\psi) - pA'(\psi) < 0.$$

Now, particularly for $\psi$ initially small, we may have

$$\frac{d\Pi_1}{d\psi} > 0.$$

(for example if $A(0) = 0, A'(0) > 0$).

### 2.2.2  Complementarity between openness and freedom: diversification effects

Now, we reintroduce control rights considerations into the analysis. A first virtue of openness in the context of the ADS model, is that it makes it easier for stage $i+1$ researcher to access ideas from stage $i$ research.At first sight, this should enhance research equally at all stages. However the argument can be made that it should enhance research to a larger extent in more upstream stages. Thus, for example, suppose that moving up to stage $i+1$ requires: (i)

either that some fixed (training) cost $K$ be spent while still working on stage $i$; (ii) or that there be a successful match between a researcher and the research idea $I_i$. The probability $\psi$ of a successful match, reflects the degree of openness in research at stage $i$.

For given $\Pi_{i+1}$, training is more attractive in the private sector than in academia, first because for given $i$ training always pays out in private sector research whereas it only pays out with probability $\alpha$ in academia, and second because paying the fixed cost $K$ is more likely to be profitable later than earlier on the line, and later stages are typically those that are operated by the private sector.

More formally, using the fact that $\Pi_{i+1}$ is increasing in $i$, there exists an integer $i_a$ such that:

$$\alpha p \Pi_{i+1} - w_a - K > \alpha p \psi \Pi_{i+1} - w_a$$

whenever $i > i_a$, and similarly there exists an integer $i_p$ such that:

$$p \Pi_{i+1} - w_p - K > p \psi \Pi_{i+1} - w_p$$

whenever $i > i_p$, and it is immediate to see that $i_a > i_p$.

Thinking of openness as increasing $\psi$, we find that more openness is of greater value in academia as academia relies more on search and matching than the private sector. Moreover, this suggests: first, that increased openness should foster the creation of new lines as these result typically from the alternative strategies pursued by free researchers; second, more openness should benefit new research institutions more than it benefits old institutions: namely, old institutions can take advantage of economies of scale or learning by doing in training, which in turn reduces the fixed cost $K$ they incur if they choose training over matching.[3]

*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.

---

[3] That more openness should foster the creation of new lines, also follows from an additional consideration, which is that openness favors cross-fertilizing 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 $\alpha$ 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 $\varphi$, 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 $\varphi$. Openness increases the net present value of a research line operated in academia, from: $\alpha p \Pi_{i+1} - w_a$ to $[\alpha + (1-\alpha)\varphi] p \Pi_{i+1} - w_a$, with $i \in \{0, 1\}$. However, openness has no value in the private sector, since researchers there are anyway forced to work on the practical strategy. With or without openness, the net present value of a line operated in the private sector, is $p\Pi_{i+1} - w_p$.

## 2.3 Testable predictions

The above discussion suggests that, particularly when researchers enjoy freedom, providing greater access to critical inputs for follow-on innovation, greater openness should enhance the total flow of knowledge building on materials that have become more open and accessible. This prediction, which is of course very intuitive, accords with a recent study estimating the positive impact of Biological Resource Centers in making key research materials available to researchers (Furman & Stern 2008). A second prediction is that the causal impact of greater openness should be more long-term because 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. This characteristic of knowledge production is well recognized by studies on the intellectual origins of critical innovations (Comroe & Dripps 1972) as well as by the analysis of the long tail of forward citations following important research articles (Garfield 1979).

But perhaps the most important predictions relate to the types of research and researchers most likely to be impacted by an "openness shock" in a world where researchers have control rights on their research activities.[4] Here 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 openness should boost basic research most when access costs are initially high and/or when control rights considerations are not first order.

# 3  Empirical framework

Across a wide variety of settings, recent decades have brought dramatic changes in the openness of key research inputs. The open source software movement, for example, has moved many programmers into an institutional setting fostering openness and sharing, although subject to wide variations in the degree of openness of the source code (Lerner & Tirole 2003; OMahony 2003). In other diverse areas of knowledge work scholars have examined the institutional arrangements (both legal and informal) supporting openness and exchange (Oliar & Sprigman 2008; Fauchart & Von Hippel 2007). Few areas, however, have been subject to more dramatic shifts in openness than the life sciences – our chosen empirical setting.

---

[4] 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.

Openness of research tools in the life sciences was dramatically shaped by the 1980 Bayh-Dole Act giving U.S. universities a broad mandate to patent and commercialize research from all disciplines (Mowery et al., 2004) and, by the late 1980s, reflected in a significant rise in patenting across universities campuses particularly the life sciences (Owen-Smith & Powell, 2004). As a result, key scientific discoveries in the life sciences, previously part of the public commons (David, 2003) were now published and claimed in patents: gene sequences (Huang & Murray 2008), biotechnology tools (Murray & Stern 2007), methods for tissue engineering (Murray 2002). This transformation sparked considerable debate over the impact of limits to openness (that potentially) accompany intellectual property rights (Heller & Eisenberg 1998; Walsh et al. 2002, 2003, 2005; Murray & Stern 2007; Heller 2008).

While IP rights can be used to reduce the openness of their key innovation inputs their owners have considerable discretion over the access terms imposed on follow-on innovators. A range of strategies exist: owners can strictly enforce their IP rights over a wide range of follow-on innovators on licensing terms they design. At the other extreme, they can commit to a non-enforcement strategy (as is the case with IBMs commitment to make 500 software patents openly available). Alternatively, owners can triage their rights differently among different follow-on innovators. In the case of research tools and materials in the life sciences, patent owners have used a variety of approaches. The case at the core of our empirical analysis highlights the licensing choices made by DuPont over patented methods to produce engineered research mice (also covering the mice themselves).

## 3.1   Openness in genetically engineered mice

With their genetic likeness to humans (the mouse and human genomes have about a 99% similarity) mice are central in the study of cancer and other human diseases (Boguski, 2002). Throughout the twentieth century, researchers in mouse genetics relied on "spontaneous mice' to for their disease studies: sick animals with specific recognizable symptoms. A dramatic breakthrough came in the 1980s when researchers combined advances in molecular biology with advances in embryonic stem cells to develop techniques for precisely engineering diseased mice. A leading member of the research community described the events that ended the classical period in mouse genetics and ushered in this important general-purpose technology: "Then" he began, "at the end of 1980, in a period of a few months, an entirely new era in mouse genetics began, with the creation of the first transgenic mice, initiated by the abrupt and then continuing entry of molecular biological techniques into what had, until then, been a classical genetic system (Paigan, 2003)".

To create an engineered mouse scientists mastered a complex process; first injecting foreign DNA into mouse eggs, then transplanting the eggs into female mice, and, if successful, observing the incorporation of the genes into the offspring (Ruddle et al. 1980, Brinster et al. 1981, Constantini & Lacy 1981, Wagner et al. 1981a, Wagner et al. 1981b). Beyond the basic techniques, sci-

entists developed more precise methods to target the insertion and expression of disease genes including the three methods we include in our analysis: Cre-lox technology controlling gene switching in tissues (Sauer et al., 1988), Oncomouse methods inserting cancer-producing (onco) genes (Stewart et al., 1984), and "Knock-out' methods deleting specific genes (Doetschman et al., 1987; Thomas & Capecchi, 1987). This suite of research tools allowed for control of the types of (disease) genes inserted into the mouse and control of the tissues where they were expressed (switched on). Their general importance was recognized in the 2007 Nobel Prize in Physiology.

The production and breeding of research mice (spontaneous or engineered) is complex, costly and time consuming. Recognizing these factors as a key limitation to the growth of their field, researchers invested in a range of scientific institutions to foster open exchange norms for research mice, facilitate their breeding and promote standardization (Radner 2001). In 1933, scientists built the Jackson Laboratories (JAX), a research organization to pioneer mouse breeding, arbitrate mouse nomenclature and standardize production procedures and serve as a repository for many strains of mice. In its role as a repository, JAX circulated a large variety of mouse strains at low cost thus enabling openness, rapid follow-on research, and low cost exchange. One leading mouse researcher looking on his mouse breeding experience prior to JAX commented on the pressures and costs of breeding and sharing his research mice: "When it was impossible to fill requests for the mice, there were grumblings that Strong was uncooperative. A few even complained that I was trying to restrict scientific material for my own selfish use. These charges were never justified. Few people realized that the inbreeds had been created in the first place as a means of opening my own scientific line of inquiry into the cancer problem. I was glad to share the mice, but I had no intention of abandoning my career in cancer research to become a supplier of laboratory animals for others (Strong 1978)".

Breakthroughs in mouse engineering increased the challenges of producing and sharing mice. Follow-on researchers wanting to use a mouse developed by another scientist could replicate their engineering methods but this required diverse knowledge – stem cell biology, molecular biology and embryology etc. Alternatively they could engage in open exchange, following the norms followed by previous generations of mouse geneticists – either informally from one researcher to another or via JAX. At first, JAX hoped to step in, extending its role in the open exchange of spontaneous mice to all engineered mice. However, IP rights on Cre-lox and Onco mice, and the licensing terms imposed by DuPont, placed severe limits on openness for all forms of exchange among researchers (in industry and academia). DuPonts control came because the new engineered mice methods were patented around the world, most notably, two types of mouse engineering methods (and the mice they generated) that are the subject of our analysis: Cre-lox and Oncomouse.

The Cre-lox technology was developed by Brian Sauer, a researcher in DuPont's life science R&D Group (Sauer et al. 1988). By inserting a particular gene "scissors' on both sides a disease gene of interest, and by incorporating a tissue-specific switch for these scissors, the Cre-lox technique produced engineered mice

with target genes turned on or off in a specific tissue. Filed in 1987 and granted to DuPont in 1990, the patent (4,959,317) covered the use of Cre-lox technology and any Cre-lox mouse. The Oncomouse patent was owned by Harvard University and covered techniques (and mice) developed Phil Leder, a professor at Harvard Medical School. The Onco patent was filed in 1984 just before the researchers submitted their Oncomouse-paper to leading journal Cell (Stewart et al., 1984). It claimed transgenic methods to insert cancer-producing onco genes into mice. The patent (4,736,866) was granted in 1988 with sweeping (And controversial) claims covering all transgenic animals (including but not limited to mice) with an oncogene (or other gene) inserted. Over the years, the breadth of this patent has been the subject of dispute and varying legal interpretation, but even the narrowest definition of the claims transgenic mice engineered to include cancer-producing genes. Since DuPont had provided unrestricted funding to Leder's laboratory – in return for right of first refusal on any patentable results – the firm received an exclusive license to the Oncomouse patent.

Throughout the 1990s, DuPont used its IP rights to control the openness of all Cre-lox and Onco-mice –whether made by public or private sector researchers. In particular, their licenses restricted the open exchange of mice among scientists (public or private). If, for example, a scientist developed an Oncomouse (or a Cre-lox mouse) then under the DuPont license they could only share it with another scientist if they complied with four terms: both parties signed the license, paid a fee, used a formal Material Transfer Agreement (contract), committed to make annual disclosures to DuPont regarding their experimental progress and granted DuPont reach through rights on any follow-on commercial applications. Their alternative was to engage in the time consuming and costly replication of the mice – although this also violated the patent.

Limits to openness caused widespread discontent among the academic community (the predominant initial follow-on innovators). They objected to a for-profit corporation using its patent rights (either as an owner or licensee) to try and change the behavior and rights to openness traditionally afforded academics (Murray, 2008). Over an eight year period (from 1990 to 1998), scientists engaged in various protests. Some shared mice informally (and against the advice of their universities). Several informal attempts among scientists to initiate to patent invalidation proceedings were never successful. At a conference in 1992 Dr. Ken Paigan, then director of JAX, announced he would make Oncomice openly available without a license, directly contravening DuPont's IP rights. While some took advantage of this opportunity, others were wary of the repercussions.

On July 1 1998, after considerable pressure from the academic community, NIH Director Nobel Laureate Harold Varmus announced a Cre-lox Memorandum of Understanding (MoU) between DuPont, the Jackson Laboratories (JAX), and the National Institutes of Health (NIH) greatly increasing the openness of Cre-lox mice for academic researchers. It allowed JAX or university researchers to distribute and share Cre-lox mice with a simple license. Before 1998, mice embodying the cre-lox technology could not be shared without a costly and restrictive license from DuPont. Post-1998 Cre-lox mice became

available for all researchers in non-profit institutions for research requiring only a Material Transfer Agreement (contract) and an institution-level license. A year later, on July 1 1999, a similar Oncomouse MoU was signed. Like the Cre-lox MoU, it applied retrospectively to all Oncomice developed prior to the MoU (as well as later mice). Taken together, the Cre-lox and Onco agreements represent dramatic shifts in openness of important research tools, and provide an ideal experimental window into the impact of openness on innovation. In formal terms, the Oncomouse embodied a similarly dramatic shift in openness, however in practice there are salient differences in the pre- and post- MoU periods that allow us to use these two openness shocks to disentangle the various predictions in our theoretical set up. Specifically, in the case of Cre-lox the MoU had the effect of increasing access and lowering the threat of expropriation of downstream research. In comparison, in the case on the Onco MoU, JAX had already made an informal commitment to openness prior to the MoU. Thus the post shock period is characterized by the same shock lowering ex post expropriation but with a much more tempered shock to open access.

## 3.2 Empirical strategy

The Cre-lox/Onco MoUs represent openness shocks for key scientific inputs exogenous to the vast majority of scientists working on mouse genetics and using engineered mice as research tools. As such, they solve a number challenges associated with evaluating the causal impact of openness on follow-on scientific research productivity. More specifically, if we used an approach comparing follow-on innovation based on knowledge produced under varying conditions of openness it would be impossible to measure the counterfactual – follow-on innovation building on the same knowledge under different openness conditions.

Take as an example, a finding that knowledge in academic publications that is more openly available generates more follow-on knowledge. We confront several possible explanations: First, publications associated with more open knowledge may simply be different in quality (higher or lower) compared to publications whose knowledge was less open. Second, knowledge associated with more openness may be different from less open knowledge and their papers might thus exhibit different patterns of follow-on research lines. For example, these papers may be more useful, more relevant or better understood. More precise causal identification must rely on some sort of experiment in openness for which the shift to more (or less) openness is exogenous both to the initial production of the knowledge and to its initial incorporation into follow-on research lines. The Cre-lox and Onco MoUs (and the comparison between them) provide such an institutional "shock' allowing us to use a differences-in-differences econometric framework within which to estimate differences in the pre- and post- shock rate of follow-on innovation, an approach that follows recent work analyzing the institutional foundations of knowledge accumulation (Furman & Stern 2006, Murray & Stern, 2007, Huang & Murray 2008, Rysman & Simcoe 2008).

Follow-on innovation is captured by taking advantage of another characteristic of the institutional environment of scientific research – citation practices.

Scientists building on the research of prior researchers acknowledge the follow-on nature of their contribution to the research line by citing the prior articles (Hagstrom, 1965; Merton 1973). While patent citations have been more widely used to trace follow-on innovation (Jaffe & Trajtenberg, 1996), citations in scientific articles to prior scientific articles are another important trace of the impact of a new discovery on follow-on research (de Solla Price, 1976; Garfield 1979; Cole 2000). We therefore take a series of scientific articles describing the production and analysis of spontaneous and engineered mice (what we refer to as mice-article papers) and trace the forward citations to these papers in other scientific papers. As the Cre-lox and Onco MoU openness shock impact some mice-article papers (those associated with Cre-lox and Onco mice-articles) and not others, and the MoUs take place with a substantial delay after the mice-article papers are published, we can exploit the timing of the shock and the non-shock mice-articles to identify the impact of the openness shocks on follow-on innovation. It should be noted that while similar in language, the two shocks may have a slightly different impact on follow-on innovation because early in the 1990s, JAX announced its willingness to contravene the strict DuPont licensing terms and distribute Onco mice, albeit without the agreement of researchers' universities. No such actions took place for Cre-lox mice. By measuring citations to Cre-lox and Onco mouse-article papers both pre and post the openness MoUs (and by measuring the citations to mouse-paper articles unaffected by the MoUs) we can separately identify the casual impact of both the openness agreements.

Of course this analysis depends upon the extent to which the Cre-lox and Onco MoU shocks to openness are 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 Onco shock were widely unanticipated in their timing and terms by the scientific community (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. The right-hand side variable in our regressions is the shock (which might be thought of as simply a "cost-shift") whereas the left-hand side variables are the forward citations which reflect the "realized demand' for prior knowledge by follow-up researchers in the pre and post shock periods. While the cost-shift may be endogenous to incentives of agents on the supply side (DuPont, Jackson, NIH), it is econometrically exogenous from the perspective of follow-on researchers. More than simply a policy announcement, or even an agreement that ratified behavior already taking place, the MoUs directly changed openness of a set of key research inputs.

## 3.3 Estimation equations

Our estimation approach uses an annual count of forward citations to a given mouse-article paper as the dependent variable and builds on recent work using citation analysis to examine the impact of institutional shocks on follow-on innovation (Furman & Stern, 2006; Murray & Stern, 2007). As a starting point,

we use a negative binomial functional form to account for the skewed nature of citation count data. Given the heterogeneity among scientific research articles and the nonlinear evolution of citation patterns over the time elapsed since publication, and the year of publication, we also include article, age and calendar year fixed effects and use a conditional fixed effects estimator to address the incidental parameters problem (Hausman, Hall & Griliches, 1984).

To identify the impact of the MoU we use the $PostOverallShock$ variable (set to one for each article-year in which a particular mouse-paper article is subject to an openness MoU). To allow for a gap between the MoU openness shock to the scientific environment and an impact on observed publication outputs, we also use the $OverallWindow$ variable which represents a short "window period' before the $PostOverallShock$ treatment period commences. We first estimate the effect of these explanatory variables on the overall level of follow-on research activity ($Citations_{jt}$) as follows:

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

where $\varepsilon_{jt}$ is the error term, $\gamma_j$ is the article fixed effect, $\beta_t$ the year fixed effect, and $\delta_{t-PubYear}$ is the article's age fixed effect.

This specification tests for the impact of the MoUs by calculating how the citation rate for a mouse-article changes after the relevant MoU, accounting for fixed differences in the citation rate across articles and relative to the non-parametric trend in citation rates for the non-treated mouse-article papers.

In the next step of our analysis, we decompose the $OverallShock$ into the two distinct experiments in our setting: the $PostCreLoxShock$ and the $PostOncoShock$ variables. In both instances we also include the corresponding window variables, $CreLoxWindow$ and $OncoWindow$, respectively. The $CreLoxWindow_{jt}$ dummy variable is equal to one if the article is subject to the Cre-Lox shock (i.e. the mouse-paper article is associated with a Cre-lox mouse), and the citation year is 1998 or 1999 (i.e. a two year window after the Cre-lox MoU). Alternatively, the $OncoWindow_{jt}$ dummy variable is equal to one if the article is subject to the Onco shock (i.e. the mouse-paper article is associated with an Onco mouse) and the citation year is 1999 or 2000 (i.e. a two year window after the Onco MoU). We again use a full set of article, age and calendar year fixed effects. We thus estimate the equation:

$$
\begin{aligned}
&Citations_{jt} \\
=\quad &f(\varepsilon_{it}; \gamma_j + \beta_t + \delta_{t-PubYear} \\
&+\Psi_{CRE_0} CreLoxWindow_{jt} + \Psi_{CRE_1} PostCreLoxShock_{jt} \\
&+\Psi_{ONCO_0} OncoWindow_{jt} + \Psi_{ONCO_1} PostOncoShock_{jt},
\end{aligned}
$$

where the coefficients $\Psi_{CRE_1}$ and $\Psi_{ONCO_1}$ measure the respective equilibrium impacts of the Cre-Lox and Onco shifts in openness on the level of follow-on

research activity.

As the predictions of our model clearly suggest, we expect to see a number of effects regarding the changing nature of follow-on innovation in the *PostCreLoxShock* and *PostOncoShock* periods. We therefore develop a specification that allows us to disentangle the shocks on different types of forward citations using a two-equation system that separates the full count of *Citations* for each citation-year into two (mutually exclusive) types and estimates the effects of our openness shock variables separately for each type. Consider for example the contrast between follow-on publications in a given year by new authors versus those who have never cited a particular mouse-article paper before.

$$NewAuthorCitations_{jt}$$
$$= f(\varepsilon_{jt}; \gamma_j + \alpha_{NEW-OLD}t + \beta_t + \delta_{t-PubYear}^{NEW}$$
$$+ \Psi_{CRE_0}^{NEW} CreLoxWindow_{jt} + \Psi_{CRE_1}^{NEW} PostCreLoxShock_{jt}$$
$$+ \Psi_{ONCO_0}^{NEW} OncoWindow_{jt} + \Psi_{ONCO_1}^{NEW} PostOncoShock_{jt},$$

and

$$OldAuthorCitations_{jt}$$
$$= f(\varepsilon_{jt}; \gamma_j + \beta_t + \delta_{t-PubYear}^{OLD}$$
$$+ \Psi_{CRE_0}^{OLD} CreLoxWindow_{jt} + \Psi_{CRE_1}^{OLD} PostCreLoxShock_{jt}$$
$$+ \Psi_{ONCO_0}^{OLD} OncoWindow_{jt} + \Psi_{ONCO_1}^{OLD} PostOncoShock_{jt},$$

The two equations are estimated jointly, with the following conditions. First, the article fixed effects are set to be identical in both regressions. Next, the calendar year fixed effects may differ only by a fixed annual rate. Finally, the age fixed effects are independent across the two regressions. This allows for different growth rates over time for public and private citations, and for differences in the flow of public and private citations received over an article's lifetime. With this setup, we then test whether, in line with the theory, $\Psi_{CRE_1}^{NEW}$ and $\Psi_{ONCO_1}^{NEW}$ are significantly larger than $\Psi_{CRE_1}^{OLD}$ and $\Psi_{ONCO_1}^{OLD}$, respectively. We develop similar specifications for several citation margins that capture the notion of diversity across research lines described in our theory: new versus old institutions, new versus old key words, and new versus old journals. We also explore research along a given research line comparing citations in applied versus basic journals.

## 4 Empirical data

### 4.1 Data and sampling

The data for this study is based on 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 in what we refer to as a mouse-article. Of this large population, we focus only on research mice publish in mouse-articles in the period from 1992 up until the first MoU in 1998. We then sample on four mouse engineering technologies (as defined by MGI). First, we include all Cre-lox mouse-articles, all of which are impacted by the 1998 Cre-lox MoU. The second group includes all Onco mouse-articles from the period; these are all impacted by the 1999 Onco MoU. The third and fourth groups represent control mouse-articles: Other Engineered mouse-articles describe new gene knockout mice not subject to the MoUs while the fourth group is the Spontaneous mouse-articles whose production relies upon random mutation rather than engineered genetic manipulation. The openness of the Spontaneous mice follows the broad sharing norms of the mouse genetics. In total, our sample includes 2638 novel mice linked to 2223 unique mouse-article papers. The breakdown is as follows: 52 Cre-lox mice, 160 Onco mice, 2171 Other Engineered mice, and 255 Spontaneous mice.

For all 2223 mouse-articles we obtained information on technology and publication year from MGI and PubMed. We then used Thomson ISI Web of Science to collect the full set of follow-on (forward) citations in academic journals from the year following publication through to the end of 2006 (a total of 525,865 citations). For each citation, we also collected 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 then aggregated by combining all the citations received by a given mouse-article in a particular year into 27,442 citation-year observations. We also then developed a set of mutually exclusive categorical variables and coded our citations accordingly. This allowed us group the citations in each citation-year for any mouse-article into two mutually exclusive citation-year observations giving us 54,884 citation-year observations in each case. This allows us to examine the impact of the MoU openness shocks on the two margins in each case.

Following our theoretical predictions, we focus first on margins capturing the predicted expansion of horizontal experimentation across new lines. We consider measures both of the diversity of researchers in follow-on innovation - new researchers and new institutional affiliations and of the diversity of research - new key words and new journals. Consider, for example, the case of new key words. We consider a key word to be new if it has never been used in a prior year to categorize a citation to particular mouse-article and old otherwise. This allows us to capture changes in the research landscape. Overall, we generate four such 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. This measure is grounded in the observation that the authorship convention in the life sciences places the faculty controlling the research last.

ii. New/Old Institution: defined as new if any address in 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 had never before appeared in the key word list of in the 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 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), using a categorization of basic versus applied journal categorization. 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). (It is worth noting that in this schema, multidisciplinary journals are classified as "basic" because this adds a conservative bias against finding an increase in basic research compared to applied research.)

Taken together these measures allow us to explore the more detailed implications of changes in openness for both horizontal and vertical follow-on research, well beyond a levels effect. Using the citation margins, we can investigate the hypothesis that changes in openness create new more diverse lines of research, pursued by a more diverse range of scientists. We also investigate were along the research line (from basic to applied) the additional research is taking place, examining the prediction that by lowering the hold-up of researchers from IP on Cre-lox and Onco mice those who already had access to research mice (researchers using Onco mice) would move downstream to more applied research. In contrast, for those without any access – such as researchers hoping to use Cre-lox mice – the main effect of lower IP would be to focus effort on basic research (at least initially). Not only does this allow us to test the rich predictions of our theoretical model but also link the importance of openness to macroeconomic models where growth is driven by technological innovation.

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.

## 4.2    Variables and summary statistics

Our empirical analysis focuses on measuring follow-on innovation through citation counts to the 2223 mouse-articles in our core data set. 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 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 the annual citation count for any mouse-article into categorical margins of interest. As outlined above, to measure diversity of citing authors, we construct the 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 a similar logic, and to capture vertical shifts in research along particular research lines, we define $BasicCitations_{jt}$ and $AppliedCitations_{jt}$, measuring 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*

# 5 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 down-stream 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 Other Engineered mice and Spontaneous mice unaffected by the MoUs and in the pre- and post- periods 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 then 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 v. 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 report the coefficient in the form of an incident rate ratio (IRR) as well as the estimated coefficient. We focus on the IRR in our presentation because it is easily interpreted: it 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.

## 5.1  Impact of Openness Shocks on Total Annual Forward Citations

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 the coefficient on $PostOverallShock$ is significant. On average, mouse-articles affected by the shocks (Cre-lox and Onco mouse-articles) received an

additional 24% 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 $PostOverallShock_{ShortRun}$ and $PostOverallShock_{LongRun}$ 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 17% increase through 2003 and a 34.5% 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 increase of 32% in citations for Cre-lox mouse-articles, compared to only a 21% for the $PostOncoShock$ variable. The difference, we interpret, as reflecting both the fact that there was some Oncomouse sharing and access prior to the Onco MoU and to the fact that the Cre-lox agreement was the first one signed by DuPont, and was therefore a greater surprise to the research community. The statistical significance of the $OncoWindow$ variable also supports our latter interpretation. Note that the main Cre-lox and Onco effects are significant at the 1% level under this specification.

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 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 increase 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 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*

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. When dividing citations into such mutually exclusive categories, such as basic and applied citations, we allow for separate age trends for the two citation categories; however, we impose common publication-year and individual article fixed effects.

## 5.2 Impact of Openness Shocks on Horizontal Exploration Across Research Lines

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 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 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. We find a 28% increase in citations by new authors, compared to only an 11% increase in citations by old authors (with the difference of the coefficients significant at the 1% level). We find similar results when dividing the overall openness shock into short-run and long-run dynamics (5-2a and 5-2b): a 22% short-run increase to new author citations compared with a 2% increase for old authors, and 40% versus 24% 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.

When we separately evaluate the Cre-lox and Onco shocks on new and existing authors (5-3a) and (5-3b), we find that the Cre-lox openness shock leads to a 44% increase in new last-author citations, but only a marginally significant 14% increase in old last-author citations. Similarly, the Onco shock leads to a 22% increase in new-author citations, but only a 13% increase in old-author citations. The short-run and long-run effects follow a similar pattern to the Overall Openness Shock dynamics: the $PostCreLoxShock$ has a 33% short-run increase and a 76% long-run increase in new-author citations, but an insignificant 0.2%

short-run rise and a significant 40% long-run rise in old-author citations. The *PostOncoShock* similarly leads to a 21% short-run increase and a 25% long-run increase in new citations, but only a marginally significant 8% short-run rise and a significant 19% long-run rise in old citations. Our results are robust to an alternative measure of authorship that uses reprint authors to designate new versus old (repeated) researcher participation.

In the final set of specifications in Table 5, 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 in 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 22.5% compared to 16.3% 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, particulary 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 26% for new key words and -1.5% (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 36% in the long run (both are significant at the 1% level). The old key word impact is small: -7% in the short run compared to 6% in the long run. 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 36% while the Onco shock is only 20%

(although both are significant at the 1% level) 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 26% increase (significant at the 1% level) in citations from new journals, and only a 13% increase in citations from old journals. We further investigate the impact of openness in (7-2a) and (7-2b). For citations from new journals, the Cre-lox shock increased *NewJournalCitations* by 42% with the *OldJournalCitations* increasing by 27%. The effect of the Onco shock is more limited by in the same direction: a 24% increase to citations in new journals, versus an 11% increase to citations in old journals.

*TABLE 7 HERE*

## 5.3 Impact of Openness Shocks on Vertical Exploitation along Research lines

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 a 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 24% during the post-shock period; at the same time, the *AppliedCitations* variable experiences 19.6% 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 108% increase in basic citations during the post-shock period, but has negligible effects on applied-research citation flows, prompting significant uptake of basic research in the area. 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.

<center>*TABLE 8 HERE*</center>

These results are further reenforced 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 92% increase in basic citations in the short ruun and a dramatic 253% increase over the next three years (through 2006). There is no significant impact on applied citations in either the short run or 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.

# 6  Conclusion

Academia has two central features: scientific freedom and openness. For researchers working within academia, the ability to control their research agenda choices and choose the ways in which they allow others to build on their research discoveries is critical (Stern 2004). In prior work Aghion, Dewatripont and Stein (2007) developed a model emphasizing the economic foundations of scientific freedom as being grounded in the granting of control rights to researchers (rather than funders). This approach emphasizes the defining characteristic of freedom within academia, rather than the more traditional view of academia as a setting in which research takes place that, on account of weak IP rights, would be subject to considerable under-investment in the private sector. Nonetheless, IP rights are still salient in academia and have been the subject of considerable debate, particularly with the proliferation of patents on key upstream innovations, include the Oncomouse and Cre-lox but also human genes (Jensen & Murray 2005), recombinant DNA techniques and even human

<center>26</center>

embryonic stem cells. As we have shown, the control rights approach provides a powerful framework within which to reexamine the role of IP rights in academia. Because the framework is grounded in the recognition that research takes place through step-by-step innovation, it is possible to examine the ways in which changing IP rights over the very early stages of research will change downstream exploitation – a view of IP rights consistent with theories that examine a single research line (Scotchmer 1996). More importantly, because the model allows for the possibility of multiple research lines, it also highlights an important, and previously overlooked, implication of IP rights over upstream knowledge. Within the control rights framework, the impact of openness of upstream research tools raises the incentives for additional upstream basic research by encouraging the establishment of entirely new research directions. In other words, within academia, restrictions on scientific openness (such as those created through formal intellectual property such as patents) limits the level of diversity and experimentation of basic research itself.

Overall we find, perhaps not surprisingly, that there is an increased overall level of follow-on research taking place after the NIH-DuPont-JAX 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. Our findings are consistent with the two distinct implications of our framework for an increase in openness. First, we have 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 find 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 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.

29

[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 Year$_j$ | Year in which article $j$ is published | PM |
| # Authors$_j$ | Count of the number of authors of Article $j$ | PM |
| Total Citations$_j$ | # 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 Year$_{jt}$ | 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 Shock$_{jt}$ | 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 Window$_{jt}$ | 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 Shock$_{jt}$ | 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 Window$_{jt}$ | 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 Shock$_{jt}$ | 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 Window$_{jt}$ | 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 Year$_j$ | 2223 | 1995.35 | 2.83 | 1983 | 1998 |
| # Authors$_j$ | 2223 | 7.034188 | 3.41921 | 1 | 34 |
| Total Citations$_j$ | 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 | 27442 | 0.0482 | 0.2143 | 0 |
| Overall Window$_{jt}$ | 27442 | 27442 | 0.0147 | 0.1204 | 0 |
| Post Crelox Shock$_{jt}$ | 27442 | 27442 | 0.0133 | 0.1144 | 0 |
| Crelox Window$_{jt}$ | 27442 | 27442 | 0.0031 | 0.0552 | 0 |
| Post Onco Shock$_{jt}$ | 27442 | 27442 | 0.0350 | 0.1837 | 0 |
| Onco Window$_{jt}$ | 27442 | 27442 | 0.0117 | 0.1074 | 0 |

# TABLE 3: SUMMARY STATISTICS BY MOUSE TECHNOLOGY

| VARIABLE | N | MOUSE TECHNOLOGY | | | |
|---|---|---|---|---|---|
| | | 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] | | |
|---|---|---|---|
| | **(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 | **[1.237]*** | | |
| Post Overall Shock Short-run | | **[1.1742]*** | |
| Post Overall Shock Long-run | | **[1.3451]*** | |
| Post Cre-lox Shock | | | **[1.3257]*** |
| Post Onco Shock | | | **[1.2117 ]*** |
| *Window* - Overall - Cre - Onco | **[1.136]*** - - | **[1.1397]*** - - | - **[1.1013]** **[1.1638]*** |
| *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%

**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] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **(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 | [1.2803]*** | [1.1059]*** | | | | | [1.2247]*** | [1.1628]*** |
| Post Overall Shock Short-run | | | | | [1.2150]*** | [1.0168] | | |
| Post Overall Shock Long-run | | | | | [1.3963]*** | [1.2351]*** | | |
| Post Cre-lox Shock | | | [1.4441]*** | [1.1414]* | | | | |
| Post Onco Shock | | | [1.2209]*** | [1.1286]*** | | | | |
| *Parametric Restrictions* | | | | | | | | |
| Separate Age FEs = 0 | | | | | | | | |
| Common Year FEs = 0 | | | | | | | | |
| Log-likelihood | -104637.73 | -104637.73 | -104627.01 | -104627.01 | -104616.36 | -104616.36 | -138282.49 | -138282.49 |
| # of Observations | 26406 | 26406 | 26406 | 26406 | 26406 | 26406 | 26406 | 26406 |

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] | | | | | |
|---|---|---|---|---|---|---|
| | **(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 | [1.2605]*** | [0.9850] | | | | |
| Post Overall Shock Short-run | | | [1.2076]*** | [0.9339]*** | | |
| Post Overall Shock Long-run | | | [1.3608]*** | [1.0602]** | | |
| Post Cre-lox Shock | | | | | [1.3586]*** | [0.9317] |
| Post Onco Shock | | | | | [1.2001]*** | [1.0209] |
| *Parametric Restrictions* | | | | | | |
| Separate Age FEs = 0 | | | | | | |
| Common Year FEs = 0 | | | | | | |
| Log-likelihood | -217511.16 | -217511.16 | -217491.52 | -217491.52 | -217473.69 | -217473.69 |
| # of Observations | 27428 | 27428 | 27428 | 27428 | 27428 | 27428 |

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] | | | | | |
|---|---|---|---|---|---|---|
| | **(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 | [1.2594]*** | [1.1263]*** | | | | |
| Post Overall Shock Short-run | | | [1.2457]*** | [1.0383] | | |
| Post Overall Shock Long-run | | | [1.2959]*** | [1.2552]*** | | |
| Post Cre-lox Shock | | | | | [1.4206]*** | [1.2698]*** |
| Post Onco Shock | | | | | [1.2365]*** | [1.1098]*** |
| *Parametric Restrictions* | | | | | | |
| Separate Age FEs = 0 | | | | | | |
| Common Year FEs = 0 | | | | | | |
| Log-likelihood | -105698.96 | | | | | |
| # of Observations | 26448 | 26448 | 26448 | 26448 | 26448 | 26448 |

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] | | | | | |
|---|---|---|---|---|---|---|
| | **(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 | [1.2383]*** | [1.1963]*** | | | | |
| Post Cre-lox Shock | | | [2.0751]*** | [0.9246] | | |
| Post Onco Shock | | | [1.0291] | [1.5605]*** | | |
| Post Cre-lox Shock Short-run | | | | | [1.9245]*** | [0.8751] |
| Post Cre-lox Shock Long-run | | | | | [2.5289]*** | [1.0719] |
| Post Onco Shock Short-run | | | | | [1.0300] | [1.5128]*** |
| Post Onco Shock Long-run | | | | | [1.0294] | [1.6298]*** |
| *Parametric Restrictions* | | | | | | |
| Separate Age FEs = 0 | | | | | | |
| Common Year FEs = 0 | | | | | | |
| Log-likelihood | -126702.71 | -126702.71 | -126504.85 | -126504.85 | -126490.16 | -126490.16 |
| # of Observations | 27414 | 27414 | 27414 | 27414 | 27414 | 27414 |

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