Running with (CRISPR) Scissors: Tool Adoption and Team Assembly

Samantha Zyontz*
MIT Sloan School of Management

May 2019

Abstract

Research tools are essential inputs to technological progress. Yet many new tools require specialized complementary know-how to be applied effectively. Teams in any research domain face the tradeoff of either acquiring this know-how themselves or working with external tool specialists, individuals with tool know-how independent of a domain. These specialists are scarce early on and can choose domain teams to create many applications for the tool or to focus on complicated problems. Ex ante it is unclear where the match between domain teams and external tool specialists dominates. The introduction of the DNA-editing tool CRISPR enables identification of external tool specialists on research teams by exploiting natural difficulties of applying CRISPR across disease domains. Teams have a higher share of external tool specialists in difficult diseases, especially for subsequent innovations. This suggests that external tool specialists and domain teams match more often to solve complex but influential problems. As more tools like Artificial Intelligence emerge, research teams will have to also weigh the importance of their possible solutions when considering how best to attract and collaborate with external tool specialists.

* Ph.D. Candidate, MIT Sloan School of Management, Technological Innovation, Entrepreneurship, and Strategic Management. I am indebted to Scott Stern, Pierre Azoulay, Jeff Furman, Florenta Teodoridis, Michael Bikard, Danielle Li, and Tim Simcoe for their guidance and suggestions that immensely improved this paper. Appreciation also belongs to Joanne Kamens and her staff at Addgene, Aditya Kunjapur, Wes Greenblatt, and other CRISPR scientists for helping me to understand the science and institutions in this setting. I also thank the participants of the Sloan-TIES reading group, the Sloan-TIES Seminar, the REER Conference at Georgia Tech, and the NBER Productivity Lunch for their insightful comments. All errors are my own.

1 Introduction

Some of history's greatest technological advances can be attributed to research tools. Research tools include physical inputs into the process of discovery and can be inventions of a method of invention with large economic impacts across a range of domains (e.g., Griliches 1957; Walsh, Arora, and Cohen 2003; Cockburn, Henderson, and Stern 2017). For example magnification provided by microscopes allowed Antony Van Leeuwenhoek to first observe bacteria, critical to today's understanding of biology and medicine (Wills 2018). The introduction of Polymerase Chain Reaction (PCR) by Kary Mullis allowed scientists to make copies of DNA which improved the speed of diagnostic tests and revolutionized the way scientists manipulated genetic material (Rabinow 2011). The statistical software package STATA improved researcher productivity in many domains, including economics, finance, epidemiology, political science, and sociology (Pinzon 2015).

Technological progress requires domain knowledge as well as tools (e.g., Rosenberg 1982, 1994, 2009; Nelson 1981, 2003; David 1990; Bresnahan and Trajtenberg 1995; Rosenberg and Trajtenberg 2004). For example, advances in microscopes and molecular pathway knowledge lead to medical breakthroughs. Over time, the combination of new tool and domain knowledge leads to an ever-larger knowledge base from which innovation emerges (Cohen and Levinthal 1989; Weitzman 1998; Fleming 2001; Mokyr 2002; Wuchty et al. 2007; Schilling and Green 2011).

Research tools require adopters to both have access and the ability to apply the tool in a domain (Teece 1986; Scotchmer 1991; Weitzman 1996; Fleming 2001). Tools often embed sufficient know-how so that primarily access to the tool lowers research costs to innovators in the domain and lowers entry barriers to external innovators (e.g., Furman and Stern 2011; Williams 2013; Murray et al. 2016). But not all new tools embed their necessary know-how. In order to effectively apply such tools, early adopters must acquire complementary tool-specific know-how. This additional input to knowledge production is distinct from human capital in the domain and physical capital in the tool. Since the complementary know-how is crucial for the adoption of a new tool that can advance technological progress this paper explores the question: *How do early teams acquire the tool-specific know-how necessary to innovate with a new research tool?*

The amount of complementary specialized tool know-how is not binary, but rather varies along a continuum. Ex ante, newly introduced tools can require more complementary know-how in some domains than others. Over a set of these domains on the continuum, the complementary know-how can come from teams internal to a domain or external specialists in the tool not associated with the domain. Over the range of domains where there is a choice of where to acquire complementary know-how, internal domain teams face a form of the "make or buy" problem over whether to learn the complementary know-how or collaborate with external specialists (e.g., Coase 1937; Williamson 1975, 1985). For example, consider a research team's decision to use advanced Artificial Intelligence (AI). If the team does not have employees trained in the technology, it must either provide incentives for employees to learn AI themselves or it must find and pay for external specialists, either through training or hiring.

When tools are new, external tool specialists are scarce, giving them a separate choice of which domain teams to join, introducing a two-sided market matching problem (e.g., Gale and Shapley 1962; Roth 1984). External tool specialists can either choose to join teams in easier domains where they can apply the tool quickly and broadly or they can choose to join teams in difficult domains where the problems are complex and solutions are potentially more influential. The decision of where and how to work with scarce external tool specialists to acquire complementary tool know-how is one firms, mangers, and individual innovators face repeatedly as they innovate.

Given the costs and benefits of collaboration and a scarce supply of external tool specialists, it is not immediately clear which domains are likely to attract external tool specialists into effective collaborations more often. However, it is reasonable to hypothesize that team assembly to acquire complementary tool know-how varies by the difficulty of using the tool in the domain and that the ex ante complexity of the *domain* determines where external tool specialists match with domain teams most often.

One concern with empirically studying knowledge recombination using research tools is that external tool specialists can contribute to the innovation process by providing access and by sharing complementary information about how to use the new technology (Polanyi 1962; Zucker and Darby 2001; Murray 2002). However, access does not necessarily lead to the ability to use a tool (e.g., Cohen and Levinthal 1990; Ahuja & Katila 2001; Zahra and George 2002; Thompson and Zyontz 2017). External specialists' contributions to the application of a new technology apart from access

are difficult to identify empirically because access is commonly conflated with the ability to use the tool or tool and domain knowledge develop concurrently (e.g., Furman and Stern 2011; Murray et al. 2016; Teodoridis 2017).

To identify whether and where external tool specialists provide specialized complementary tool know-how to create early innovations, an ideal setting would separately identify external tool specialists. The setting would also allow access to be tracked separately from the ability to use the tool. Finally, the ex ante difficulty of applying the tool should vary across domains, and the tool should enter different domains randomly. Such an environment does not occur naturally, but the recent introduction of the DNA-editing tool, CRISPR, provides a novel and approximate setting.

CRISPR is a naturally occurring immune response in bacteria that proved to be a powerful editing tool for DNA modification in almost any organism. The CRISPR tool was first introduced between June 2012 and January 2013 when researchers from Berkeley, MIT, and Harvard demonstrated that CRISPR could be used to edit DNA in both bacteria and mammals. CRISPR is a substantial improvement over existing DNA editing tools especially for researchers working in mammalian cells (Zyontz 2016). For mammalian researchers, it represented a long-term reduction in research costs and an unanticipated increase in new opportunities across a wide range of domains.

CRISPR has incredible promise in human disease domains caused by genetic mutations, such as infections, viruses, and inherited genetic diseases. However, not every such disease received access to CRISPR at the same time, as much as researchers wanted to adopt the tool. Although CRISPR works similarly once delivered into a cell affected by a disease, certain cells are biologically more difficult to edit than others. This imposes natural delays on when CRISPR can be applied to a particular disease, since researchers must first overcome this delivery problem. The unanticipated timing of the CRISPR tool introduction to different human disease domains mitigates some of the endogeneity inherent in adoption and helps to identify the knowledge bases of the innovators responsible for the articles that use CRISPR in a disease.

Specifically, the CRISPR tool was not originally developed for any specific domain application because it was a shock to gene editing. It took almost another year for a human disease application to appear. This established a set of separate CRISPR tool specialists not associated with a disease but whose know-how would be useful in any disease domain. In the earliest years (2012 – 2016), the

primary adopters of CRISPR were academic scientists so it is possible to use historical publications from PubMed to separately identify external tool specialists from domain specialists in CRISPR.

Because biological materials are often transferred between labs, science settings can conflate the different access and complementary know-how contributions of an external tool specialist. Material Transfer Agreements (MTAs) can delay access or require co-authorships for the receipt of the materials (Walsh, Cohen, and Cho 2007; Strandberg 2010). The CRISPR setting circumvents these concerns by having a biological resource center that breaks the link between contracting for access to the tool and acquiring complementary know-how from external specialists. From 2012-2016, Addgene was the primary distributor of CRISPR to academic researchers and as a third-party, eliminated the need for scientists to sign MTAs directly with other academics. Thus any external tool specialists observed entering a new domain in this setting primarily represent value added know-how rather than the price of access.

The results show that the share of external tool specialists on new CRISPR papers in a disease is significantly larger in more difficult disease domains, suggesting that the match between external tool specialists and domain teams occurs more often in domains that focus on solving influential problems rather than the breadth of implementation. The share of external tool specialists also increases for subsequent CRISPR papers in difficult disease domains, so the effect does not attenuate immediately. The paper is the first to introduce CRISPR as a setting to empirically study how teams form to effectively overcome tool adoption barriers in know-how. It also uniquely shows that effective team composition is driven by the specific nature of the problem and the nature of tools available for innovation, not just features of management, organizational structure, or industry. As more tools emerge that require the acquisition of complementary tool know-how, like AI, research teams looking to be early adopters of such tools will have to weigh the complexity and importance of their possible solutions in considering how best to attract and collaborate with external tool specialists.

Section 2 discusses the relevant literature on knowledge inputs to innovation production and the choices faced by internal domain teams and external tool specialists. Section 3 provides details on the CRISPR setting. Section 4 outlines the identification and empirical specifications used for the analysis. Sections 5 and 6 describes the measures constructed and results. Section 7 provides a discussion of the results and concludes.

2 Tools as Inputs to Innovation Production

The recent emergence of AI can be used to innovate in a range of applications including natural language processing, image recognition, enhanced data security, and smart products like cars (Marr 2016). Using AI tools as inputs to innovation in these areas requires more than just accessing an off-the-shelf product. AI also requires adopters to acquire specialized knowledge and skills including coding, training models, building computing infrastructure, and scaling for firm-wide implementation. The amount of complementary knowledge needed varies by application area. From a firm's perspective, it can either have their employees learn the complementary AI know-how internally or can bring in external AI specialists. External specialist know-how can be useful across AI applications, but when external specialists are scarce they also get to choose the application areas in which to collaborate. For example, the initial team at Google working on the self-driving car consisted of Google-X employees with engineering and AI experience. The team could have eventually learned the complex complementary AI know-how internally, but instead Google-X collaborated with a specialist previously running the Stanford Artificial Intelligence Laboratory at Stanford University, Sebastian Thrun (Dallon 2017). However, Dr. Thrun's also had the viable choice to join the team or work on different applications since his skillset was rare.

The AI anecdote illustrates how recombining external specialized know-how with internal domain knowledge can help research teams move from access to ability to use a new tool. It also suggests that complexities in a domain may influence the match between internal domain teams and external know-how. External tool specialists may be attracted to teams doing more complex and influential work when their skillsets are in demand and scarce.

2.1 Research Tool Adoption and Innovation

Research tools are types of inputs to innovation distinct from human capital in a domain. Tools are integral to technological progress in an application domain because they often embed their own know-how, allowing users to apply the tool without understanding why it works (Mokyr 2002). Access to these tools helps to lower the costs of research to innovators in the domain and invites those outside of the domain to make new contributions (e.g., Furman and Stern 2011; Williams 2013; Murray et al. 2016; Teodoridis 2017; Furman and Teodoridis 2018).

Much of the traditional adoption literature focuses on the diffusion of products throughout their lifecycles and their role in economic growth or social returns. Related literature discusses the types of individuals who adopt these products, but neither focuses directly on the role of research tools in creating innovations. Some of the earliest work on product adoption looked at the social rate of return to hybrid corn research (Griliches 1957, 1958) and showed that although there is a high return to investment, it takes time for products to diffuse due to both availability and acceptance. Locations most in need of the new product will likely adopt it sooner, but even within an area diffusion occurs in an "S-shaped" pattern as some individuals wait to adopt. Complementary work on adopter types showed that the earliest product adopters at the beginning of the S-curve are influential in their fields, have resources to adopt, and have the ability to incorporate the new products (Rogers 1962).

The literature on general purpose technologies (GPTs) also focuses on the role of technological progress to economic growth. GPTs can be technologies such as semiconductors (e.g., Bresnahan and Trajtenberg 1995), steam engines (Rosenberg and Trajtenberg 2004), or information and communication technologies (e.g., David 1990). GPTs generally are considered enabling technologies that encourage economic growth in a large range of downstream sectors through a positive feedback loop between a GPT producer and downstream markets as each makes complementary improvements to the GPT (Bresnahan and Trajtenberg 1995). Different sectors may delay adopting the GPT depending on when it is most valuable.

Innovation as an outcome for research tool adoption is more common in empirical work that focuses on access. Access to research tools has been shown to lead to an increase in the rate and changes to the direction of innovation in a number of settings (e.g., Moser 2005; Azoulay et al 2009; Furman and Stern 2011; Williams and Sampat 2015; Murray et al. 2016; Teodoridis 2017). However, these papers tend to focus on access to tools that reduce the cost of research quickly with little tool-specific know-how needed. Further, it is assumed that the new and broader work is due to innovators using the tool they can now access. That mechanism is not assured though since the tool is rarely tied directly to the new papers or products.

Furman and Teodoridis (2018) have one example of tool know-how interacting with different domains, but even in the case of Kinect, the tool is assumed to reduce the cost of research at the time of access with little variation. Therefore it is not possible to use the difficulty of applying the

tool to understand how tool specific know-how is incorporated. Further the authors, by necessity, assume that the internal and external researchers they study are using the Kinect tool. Nagle and Teodoridis (2017) take a more direct look at the role of researchers who use Kinect and show that it is generalists who tend to bring the new tool into teams. Once again, because the costs of Kinect do not vary by domain, they cannot address how their outcomes might change as the tool is more difficult to use and the problems become more complex or influential.

This paper adds to the literature on research tool adoption and innovation by introducing a way to empirically observe innovations directly due to a tool that not only requires complementary know-how to use but the amount necessary varies by application domain. It also uniquely shows both the nature of the problem and the nature of tools available for innovation affect successful team composition, not just features of management, organizational structure, or industry.

2.2 Research Tools and Complementary Know-How

Many tools that have been historically important for technological advancement are those that eventually embed the necessary know-how in the physical product including hammers, scissors, microscopes, steam engines, automobiles, or telephones. For these, the decision to adopt is mostly rooted in access to the tool.[1] However, some tools when first introduced do not embed necessary know-how and require the user to learn complementary know-how even after obtaining access. For example, tools like early wind tunnels, the first computers, or early software packages like STATA all embedded some of the underlying knowledge of physics, computing algorithms, or statistics. However, users required additional knowledge of the tool in order to apply it to different problems. For example, the earliest versions of STATA could run multiple regression analysis directly, but time-series analysis and other more advanced statistical models still needed to be programmed by the user (Pinzon 2015).

The above might suggest that tools should be classified dichotomously – those that need no additional know-how and immediately lower learning costs versus those that need a host of complementary know-how to bring learning costs down. However, this is an oversimplification. Instead, it is possible to think of tools appearing on a continuum based on the amount of

[1] Although, at introduction, many tools embed less of their own know-how than they do after they become more routinized.

complementary know-how needed to employ the tool in a domain at a particular time. Some tools like hammers are introduced with all necessary embedded know-how so their positions on the continuum generally do not change over time. Other tools embed a greater amount of knowledge over time. Continued use of these tools bring about improvements in performance and modifications that embedded more knowledge in the tool itself. For example, as users created code for more advanced statistical models, later versions of STATA included those updates so that non-statisticians could apply the models just as easily. Thus later versions of the tool can appear in different locations on the continuum than the original. Finally, the same tool can appear in different locations on the continuum based on the application domain at a given point in time.

If tools lie on a continuum of least to most necessary complementary know-how, then there is a range of tools that require the user to learn complementary know-how as an additional input to innovation. One option for adoption of tools within this range is to delay until the tool is improved and embeds enough needed know-how, which is an aspect of adoption discussed in the adoption literature (e.g., Griliches 1957, 1958). However, our understanding of how early users adopt tools when this complementary know-how is still needed is incomplete. Understanding how early teams form to effectively adopt tools that require complementary know-how and use them in innovations provides a better idea of the choices that shape early adoption and the innovative paths that are formed from these early decisions by successful teams.

For teams already in an application domain that want to be early adopters, they can choose to learn the necessary know-how internally or they can collaborate with external tool specialists, those that acquired the tool know-how independent of a domain. However, external tool specialists are scarce in this early stage and can choose in which domains to work. Effective collaborations will only occur in domains where there are sufficient incentives for both sides.

2.3 Domain Team Choices

Over a range of domains where a newly introduced tool requires the user to learn complementary know-how, teams that want to be early adopters face a form of the "make or buy" problem (e.g., Coase 1937; Williamson 1971, 1975, 1985; Grossman and Hart 1986). Teams can invest time learning the tool know-how internally (through the team leader or another team member). Alternatively, they can choose to acquire the know-how from external tool specialists who

developed human capital in the tool independent of the domain (e.g., Cassiman and Veugelers 2006 and Grigoriou and Rothaermel 2017). However, in order to successfully use external knowledge, an organization must not only have access to the new knowledge but must also have the resources and ability to incorporate it (Cohen and Levinthal 1990).

The choice to internally learn the complementary know-how or acquire it from external tool specialists involves weighing the costs and benefits of each option. For example, learning internally has the benefit of providing the internal domain team control over its work, making it more self-sufficient for future innovations. The drawback is that learning a new tool can take time, possibly causing the team to give up a first mover advantage (Jones 2009). Collaborating instead with external specialists can reduce the time it takes to apply the tool since the specialists bring the complementary know-how with them (e.g., Arora and Gambardella 1994; Wuchty et al. 2007; Uzzi et al. 2013). However, collaborations have inherent frictions that need to be overcome before the tool can be applied so there is a risk that the collaboration could fail (Cummings and Kiesler 2007; Bikard et al. 2015). Internal domain teams will seek external tool specialists if the difference between the costs and benefits of collaborating is greater than the difference between the costs and benefits of learning internally.

2.4 External Tool Specialists Choices

Although innovation often emerges from a recombination of previous ideas (e.g., Scotchmer 1991; Fleming 2001; Kaplan and Vakili 2015), researchers need both access to a tool and the ability to incorporate specialized tool know-how (Teece 1986; Scotchmer 1991; Weitzman 1996; Fleming 2001) to successfully innovate. However, access does not necessarily lead to the ability to use a tool (e.g., Ahuja & Katila 2001; Zahra and George 2002; Thompson and Zyontz 2017). External tool specialists can contribute to the innovation process by sharing information about how to use the new technology (Polanyi 1962; Zucker and Darby 2001; Murray 2002).

When a research tool is introduced, the stock of external tool specialists is initially small. If there is a rush to use the new tool, scarce external tool specialists can choose the teams and domains where they wish to share their complementary tool know-how. If external tool specialists choose to collaborate with a team, it may be because collaborations are increasingly common and they have been shown to result in higher productivity and higher quality ideas (Adams et al. 2005; Stephan

2012; Gans and Murray 2014). By collaborating with teams in easier domains where the tool needs less complementary know-how, external tool specialists can be more productive and broadly apply the tool to more domains quickly. On the other hand, by collaborating with teams in difficult domains where the tool needs more complementary know-how, external tool specialists can work on more challenging and complex problems where the possible solutions are highly influential. The external tool specialists' choice creates a two-sided matching market for the scarce human capital in the complementary tool know-how. Internal domain teams and external tool specialists will only collaborate in domains where there is a match on both sides.

Given the costs and benefits of collaboration to internal domain teams and external tool specialists, it is not ex ante obvious which kinds of domains are likely to attract external tool specialists into collaborations more often. However, it should be the case that the ex ante difficulty of an application domain will determine where the successful matches occur. To test how early teams acquire specialized complementary tool know-how to produce innovations, the recent introduction of the DNA-editing tool, CRISPR, provides a novel setting where the knowledge bases of the earliest innovators can be separately identified using natural variation in CRISPR's entry into different domains. The next section describes CRISPR and useful factors of the setting.

3 Gene Editing with CRISPR

CRISPR provides a unique setting for exploring how innovative teams acquire specialized complementary tool know-how. First, CRISPR was an unexpected shock to gene editing researchers and the DNA-editing tool's use has exploded since its introduction. Second, access to the tool is widely available through the biological resource center Addgene, alleviating the need for material transfer agreements between labs. Third, CRISPR can be applied to many different disease domains but was not developed for a particular application. The difficulty of applying CRISPR to different diseases varies based on natural properties of the cells to be edited.[2]

[2] Further details on CRISPR and its introduction to gene editing beyond those provided in this section can be found in Zyontz (2016) and Thompson and Zyontz (2017).

3.1 Gene Editing Before CRISPR

DNA editing has led to many advances including the creation of model organisms and the modification of existing organisms since its introduction in the early 1970s. These advances allowed researchers to better understand human disease and to create useful products like pesticide resistant crops. For bacteria or other prokaryotes that have cells without nuclei, relatively easy editing techniques existed prior to CRISPR. However, for higher-order species, like mammals, even recent editing alternatives like Zinc Finger Nucleases (ZFN) and Transcription Activator-Like Effector Nucleases (TALENs) (Moscou and Bogdanove 2009; Boch, et al. 2009) are difficult and time-consuming to use. Despite this, TALENs was chosen as "Method of the Year" in 2011 (Method 2012) because of the advancements it represented. Only a few months later, CRISPR came as a surprise to researchers working in gene editing. The new CRISPR tool acts like a pair of universal DNA scissors that work across organisms and is a much more flexible option for DNA editing than any other tool, especially for complex organisms like mammals.

3.2 CRISPR

The exact purpose and function of CRISPRs, short for Clustered Regularly Interspaced Short Palindromic Repeats, were not well understood until 2007 when it was discovered that the unique DNA sequences are an adaptive part of the bacterial immune system (Barrangou et al. 2007). Bacteria use CRISPR sequences to recognize viral DNA and then use a related enzyme (often the Cas9 protein) to cut up invading viral DNA and destroy the virus.

In June 2012, Professors Jennifer Doudna and Emmanuelle Charpentier at the University of California, Berkeley first introduced a modifiable CRISPR system for DNA editing (Jinek et al. 2012). Doudna and Charpendier proved in test tubes that the CRISPR system could find and edit any DNA sequence (not just viral) using programmed guide sequences to direct the cutting enzyme to the right place in the DNA. Doudna and Charpentier's work provided proof of concept that the CRISPR system could edit organisms like bacteria. In January 2013, MIT Professor Feng Zhang and his collaborators showed that the CRISPR system could also edit mammalian cells, including human cell lines (Cong et al. 2013). This work and the related work of George Church and his colleagues at Harvard Medical School (Mali et al. 2013) demonstrated the flexibility and ease of use of the new

CRISPR tool. This was particularly welcomed in the mammalian research community, including those working on human cells, because of the enormous improvement in terms of accuracy and difficulty over previous methods.

To understand how CRISPR works, consider the find-and-replace function in a word processing program. To replace the word "absolutely" with "certainly," the user only need to put in the two strings and the program will find all instances of the string "absolutely," cut it out of the document, and replace it with the second string. CRISPR works the same way: program a DNA sequence to find the same string in the DNA of an organism, use an enzyme to cut out that string, and then use a second programmed string as a replacement in the organism's DNA. One of the main benefits to CRISPR is that it can look for the equivalent of the full word "absolutely." Previous editing technologies could only search for a short string, like the equivalent of "abs." In the find-and-replace analogy, this short string would find "absolutely" but also "abstract," making the edited document (and edited DNA) unreadable.

The enthusiasm from researchers conducting gene editing to the release of the CRISPR tools was almost immediate because of its accuracy, flexibility, and relative ease of use (Pennisi 2013; Regalado 2014). Since Doudna and Charpendier's first paper in June 2012 through December 2016, over 4,500 CRISPR-related articles were published according to the medical publication database PubMed. Patent applications mentioning CRISPR and funding for venture backed firms licensed to use CRISPR technology have also soared (Ledford 2015). By December 2016, over 3,000 patent applications published worldwide mentioned CRISPR. Funding also flowed to CRISPR commercialization efforts. The original biotech firms founded on CRISPR technology, Caribou Biosciences (Berkeley, CA), Editas Medicine (EDIT; Cambridge, MA), CRISPR Therapeutics (CRSP; Basel, Switzerland), and Intellia Therapeutics (NTLA; Cambridge, MA) collectively raised initial funding of more than $150 million. The last three all had IPOs in 2016, each currently with market capitalizations of over $1 billion.

CRISPR's effect on biological research has been profound, as geneticist John Schimenti at Cornell University noted: "I've seen two huge developments since I've been in science: CRISPR and PCR… CRISPR is impacting the life sciences in so many ways" (Ledford 2015). One of the original inventors, Jennifer Doudna, stated in a February 2015 JAMA editorial, "This discovery has triggered a veritable revolution as laboratories worldwide have begun to introduce or correct

mutations in cells and organisms with a level of ease and efficiency not previously possible." (Doudna 2015). CRISPR has already been used to create blight resistant crops (Wang et al. 2014) and "malaria-proof" mosquitoes that are genetically unable to transmit malaria (Gantz et al. 2015). The introduction of CRISPR is likely to be especially useful in medical applications since it may ultimately allow for the correction of genetic errors.[3] Currently, it is allowing researchers to build mouse and human cell disease models more easily with specific mutations that are useful for testing drugs. For example, Feng Zhang's lab has created a "Cas9 mouse" (Platt et al. 2014) that can be modified to model lung cancer. Before CRISPR, creating such a mouse model took large teams of people and a decade to complete, but this model was designed by one person with CRISPR in four months (Specter 2015).

3.3 Access to CRISPR with Addgene

Laboratories often use material transfer agreements to transfer tools from one institution to another, often delaying adoption of the tool (e.g., Mowery and Ziedonis 2007; Walsh, Cohen, and Cho 2007; Strandberg 2010). To circumvent the difficulties associated with lab-to-lab material transfers, biological resource centers, such as the American Type Culture Collection, are created to centralize the distribution process (Furman and Stern 2011). The dominant central repository for CRISPR, from the first day, is Addgene. In 2004 Melina Fan, Kenneth Fan, and Benjie Chen founded Addgene as a non-profit biological resource center for scientists to easily share tools[4] for use in biological research (Fan et al., 2005). Addgene not only stores biological tools donated by academic researchers all over the world, but also validates the materials and facilitates their distribution to other academic institutions in more than 85 different countries and counting.

When the CRISPR tool was first introduced in 2012 and 2013, Doudna, Charpendier, and Zhang donated their versions to Addgene at the time the original papers were published. As Zhang said in a talk at MIT in 2015, he gave CRISPR to Addgene for distribution because his lab "wouldn't have time to do science" if they responded to all the requests from other researchers. Indeed, to date, Addgene has sent more than 42,000 CRISPR tools from Zhang's lab to over 2,000

[3] This has already been done in non-viable human embryos (Ma et al. 2017) and may have been done in viable embryos resulting in gene edited babies as recently as November 2018 (Cyranoski 2018).

[4] The tools donated to Addgene, including CRISPR, are usually distributed as plasmids. A plasmid is a form of circular DNA that is commonly used to replicate or expand upon gene editing experiments. See https://www.addgene.org/ for available tools.

institutions (Zhang 2018). As new CRISPR tools have been developed, they have also been donated to Addgene. CRISPR orders quickly rose from 0.1% of all Addgene orders in 2012 to about 18% in 2015 (Figure 1). To date, CRISPR is one Addgene's most popular tools, making up over 20% of total orders.

[Insert Figure 1 about here]

Addgene has a price of $65 per plasmid which has remained constant since 2004. This stable low cost and consistent quality control process encourages labs to order directly from Addgene rather than make their own or attempt to get it from the original lab. Addgene alleviates the burden on the individual research labs and separates access to the tool from the original inventor.

3.4 Variation in Availability of CRISPR by Disease Domain

When the CRISPR tool was first introduced in 2012, it came as such a surprise that it was not designed with a specific application in mind. However, some of the greatest strides forward have come from mammalian gene editing particularly in human diseases (Zyontz 2016). The co-founders of CRISPR argue that the tool could be useful for the study and eventual treatment (or even cure) of most human diseases caused by to genetic mutations (Doudna and Sternberg 2017; Whitaker 2018), making CRISPR a valuable tool for research in all of these domains. Despite CRISPR's generality, natural barriers in gene editing prevent CRISPR from being available for all disease domains at the same time. Availability variation is due in part to the type of cell primarily targeted by the disease, which affects how CRISPR must be delivered to the nucleus of the cell (e.g., Regalado 2016, Stockton 2017, Kaiser 2016, Wang et al. 2016, LaFountaine et al. 2015). One of the co-founders of CRISPR, Jennifer Doudna, notes in her book, "That's not to say that it'll be easy to get CRISPR inside the cells themselves. This delivery problem is one of the greatest challenges" (Doudna and Sternberg 2017). The more complicated the disease and cell type, the more difficult it is to deliver and use CRISPR, which provides natural delays that vary by disease as researchers overcome the delivery barrier.

Some of the easiest diseases to edit with CRISPR involve cells that quickly self-replicate and can be edited *ex vivo*.[5] For example, blood cells can be easily removed from an organism, edited

[5] In *ex vivo* (exterior) gene editing, target cells are first modified outside a living organism. The edited cells are then returned to the organism as an effective treatment for the disease.

with CRISPR in a dish, and then the modified cells are put back in the organism (Regalado 2016). Because blood cells replicate easily, the new ones will replicate with the edit and eventually overtake the old damaged cells. T-cells associated with immune deficiencies can also be successfully edited *ex vivo*. Recently CRISPR was used to edit infected T-cells and eliminate HIV in mice (Stockton 2017).

More complicated diseases to edit with CRISPR involve cells that can self-replicate but may not be prime targets for *ex vivo* editing. For example, using CRISPR to study diseases in muscle tissue is more difficult than blood cells (Regalado 2016). Studies are underway to treat Duchenne muscular dystrophy (DMD) (LaFountaine et al. 2015), but edits must be made to all damaged cells, which cannot be done effectively *ex vivo*. Treating DMD requires a delivery mechanism that targets the affected cells, is large enough to deliver the CRISPR system in the organism, and does not make the individual sicker.

Some of the most difficult diseases to study and treat with CRISPR are those involving cells that do not replicate and cannot be edited *ex vivo*, such as diseases in the brain or nervous system (Regalado 2016, Kaiser 2016, LaFountaine et al. 2015). Brain cells and nerve cells are generally difficult to manipulate in a lab setting and because their interconnections matter, studies generally are conducted *in vivo*.[6] Again, this results in a delivery problem, where the engineered CRISPR tools are too large for standard delivery mechanisms. Work is being conducted to shrink the size of the cutting enzymes and to find alternative delivery mechanisms (Wang 2016), but the additional limitations have delayed research in these areas. For example, CRISPR has only appeared in Huntington's Disease publications since 2017.

There are already a number of reported possible therapeutic applications of CRISPR (LaFountaine et al. 2015) including cystic fibrosis (2013), HIV-1 (2013), sickle cell anemia (2014), Hepatitis B (2014), Duchenne muscular dystrophy (2014), HPV (2014), and a range of cancers. This suggests that applications where CRISPR is easier to use generally gain access to CRISPR sooner. Although no clinical trials with CRISPR were approved in the U.S. until 2018.

[6] In *in vivo* (interior) gene editing, target cells are modified while still inside the living organism.

4 Methodology

4.1 Using the CRISPR Setting

As discussed in Section 2, external tool specialists can add value to the innovation process by sharing information about how to use the new technology across a range of domains. The match between internal domain teams and external tool specialists to create new innovations with the tool could occur more frequently in easier domains, where less complementary tool know-how is needed, in order to apply the tool more broadly or in more difficult domains, where more complementary tool know-how is needed, where the problems are more complex but influential. However, empirically identifying where external tool experts contributed most often to the application of a new tool, beyond just access, is a complex task as tool and domain knowledge often develop concurrently or access is conflated with the ability to use the tool (e.g., Furman and Stern 2011; Murray et al. 2016; Teodoridis 2017).

In order to identify where external tool specialists appear more often to create early innovations with a tool, an ideal setting would need to have several main features. First, it must be possible to separately track external tool specialists and domain specialists within teams that generate early innovations. Second, the sharing of know-how by external tool specialists must be separable from any access to the tool they may provide in order to identify their value-added contributions. Finally, when the new tool is introduced, the difficulty of using the tool must vary across domains and the tool should randomly enter different domains to mitigate selection based on the value of the tool to the domain. Such an environment does not occur naturally, but the recent introduction of CRISPR provides a unique and approximate setting for this ideal.

As discussed in Section 3, CRISPR was an unexpected, powerful tool that had the potential to lower the costs of research and create new possibilities in gene editing. Because of its unexpected nature, gene editing scientists did not anticipate its arrival or the future advances the tool would eventually allow. CRISPR can be used in a wide range of organisms including bacteria, yeasts, plants, insects, and mammals. However, previous research has shown that scientists who conduct gene editing on organisms from different branches on the biological tree of life (e.g., bacteria versus mammal) have different uses for the tool and value CRISPR differently (Zyontz 2016). This argues

for focusing on researchers most at risk for using CRISPR, namely scientists conducting research on mammalian cells. During the earliest period of CRISPR adoption (2012-2016), these scientists are mostly academic scientists publishing in academic journals.

To ensure that different domains do not adopt CRISPR only when it is most valuable, it is necessary to find areas of mammalian research where scientists want to use CRISPR immediately, but cannot for some biological reason. Fortunately, research in almost every DNA-altering human disease would benefit from CRISPR due to its improvements over the previous tools. CRISPR co-inventor Feng Zhang supported this claim in a recent interview saying, "There are about 6,000 or more diseases that are caused by faulty genes. The hope is that we will be able to address most if not all of them" (Whitaker 2018). However, not every disease received CRISPR at the same time. The timing of CRISPR's introduction to each disease was not anticipated due to natural delays caused by the biology of the cells affected by the disease, as highlighted in Section 3.[7]

By restricting attention to human diseases caused by mutated genes, including infections, viruses, and inheritable diseases, the setting provides a way to separately identify external tool specialists and domain specialists. CRISPR was introduced first as a tool in June 2012 with no specific human disease application. Since the first human disease application occurred in 2013, there was time for scientists to gain CRISPR-specific tool knowledge outside of any particular disease domain. Applying this initial delay and the unanticipated timing of CRISPR appearing in different human diseases, it is possible to identify which authors were specialists in CRISPR prior to publishing in the disease. Figure 2 provides an example of the natural delays in CRISPR entry for selected diseases from first introduction of the tool.

[Insert Figure 2 about here]

Finally, science settings are not always ideal for identifying the contributions of external specialists beyond providing access to the tool. Biological materials are often transferred from one lab to another using Material Transfer Agreements (MTAs) that can delay access or have been known to require co-authorships for the receipt of the materials (Walsh, Cohen, and Cho 2007; Strandberg 2010). This can confound the value added by external tool specialists in innovation.

[7] Certain diseases attract more attention and funding which could mitigate delays in receiving CRISPR. However, the exact timing of CRISPR's arrival in these diseases still could not be anticipated *ex ante*.

The CRISPR setting circumvents these concerns by having a biological resource center that breaks the link between contracting for access and needing to work with external specialists to use the tool. Addgene is the primary third-party distributor of CRISPR to academic researchers that eliminates the need for scientists to sign MTAs directly with other academics. Thus any external specialists observed entering a new domain in this setting primarily represent value added knowledge rather than the price of access to CRISPR.

4.2 Empirical Specifications

To explore how early teams acquire complementary know-how to use new tools for innovation, the empirical specifications in this paper test the relationship between the share of external CRISPR tool specialists authoring CRISPR papers in a set of disease domains and the difficulty of editing the cell targeted by the disease. The focal population consists of all successfully published CRISPR articles in a set of human diseases and their authors. The level of analysis is at the disease-quarter from Q1 2013, the first quarter a CRISPR disease application appeared to Q4 2016. The vast majority of the disease quarters only contain one CRISPR-disease paper, so the share of external CRISPR specialists can be interpreted as the relative participation of external CRISPR specialists on the team for one paper (or innovation). The specifications also consider the unanticipated timing of CRISPR's entry by disease.

The main specification tests the overall impact of target cell editing difficulty on the share of external CRISPR specialists who are authors on CRISPR publications in a disease. The model controls for the quarter of publication and the amount of time since the first CRISPR paper in the disease to account for factors specific to the quarter or the trend of additional papers. The model also controls for the total number of publications in a disease quarter as a proxy for the attention and funding a disease may receive.

$$Share\ of\ External\ CRISPR\ Specialists_{it} = \beta_0 + \beta_1 Edit\ Difficulty_i + \beta_2 Total\ Disease\ Pubs_{it} + \delta_t + \delta_{age} + \varepsilon_{it}$$

Share of External CRISPR Specialists$_{it}$	=	The number of CRISPR paper authors that are external CRISPR specialists divided by the total number of CRISPR paper authors by disease domain (i) and quarter (t).
Edit Difficulty$_i$	=	1 if the target cells of the disease (i) cannot be edited *ex vivo* (*No Ex Vivo*) or do not self-replicate (*No Cell Replication*); 0 otherwise. Two separate variables.
Total Disease Pubs$_{it}$	=	The total number of papers in disease domain (i) and quarter (t).
δ_t	=	Fixed effects for the quarter (t) of publication.
δ_{age}	=	Fixed effects for the difference between the focal quarter of publication (t) and the quarter of the first CRISPR publication in a disease domain (i).
ε_{it}	=	Error term.

In this model, the coefficient of interest is β_1, which is the relationship between disease cell editing difficulty and the share of external CRISPR specialists in teams that publish CRISPR papers in a disease. A negative coefficient would support the idea that internal domain teams and external tool specialists match more often in easier domains that can help increase productivity and encourage broader use of the tool. A positive coefficient would support the idea that internal domain teams and external tool specialists match more often in difficult domains to effectively address more complex but influential problems.

The second specification tests whether the share of external CRISPR specialists increases or decreases for subsequent innovations after the first in difficult diseases. It is similar to the first specification above, but adds an interaction term and disease (i) fixed effects.

$$Share\ of\ External\ CRISPR\ Specialists_{it} = \gamma_0 + \sum_{age} \gamma_{age} Edit\ Difficulty_i * Qtr\ from\ First\ Pub_{age} + \gamma_2 Total\ Disease\ Pubs_{it} + \delta_i + \delta_t + \delta_{age} + \varepsilon_{it}$$

Where *Qtr from First Pub* is the difference between the focal quarter of the publication (t) in a disease (i) and the first CRISPR publication quarter in the disease. The coefficients of interest in this model are γ_{age} which are the changes in the share of external CRISPR specialists authoring subsequent CRISPR papers in a disease for difficult to edit target cells. Positive coefficients indicate that a higher share of external CRISPR specialists participate in additional innovations in domains where CRISPR is more difficult to use, even after controlling for time effects, age effects, disease effects, and the attractiveness of the disease.

All specifications are run initially as OLS models since the outcomes have a continuous response between 0 and 1. However, because the outcome is fractional and has some weight on 0 and 1 values, simple linear models lead to predictions outside the possible range. To mitigate this concern, all specifications are also run as Generalized Linear Models (GLM) with binomial family and logit link. The latter specification is as outlined by Papke and Wooldridge (1996) who showed that quasi-maximum likelihood estimation (QMLE) for pooled fractional response models results in robust estimators. The direction and significance of the results are similar regardless of model used, although the coefficients have different interpretations.

5 Data and Measures

5.1 Database Construction

Because CRISPR can be traced to a handful of initial papers and because of the explosion of interest in the tool, it is possible to find the entire population of academic scientists using CRISPR in human diseases, and not just a sample. The database starts with all articles and authors in the U.S. National Library of Medicine's (NLM) PubMed database from 2007-2016, providing approximately five years before CRISPR and five years after. PubMed includes the NLM's MEDLINE journal citation database and contains 28 million citations for biomedical literature

including the fields of biomedicine and health, making it a definitive source for original research in human diseases.

In order to identify the relevant disease domains used in this study the keywords assigned by the NLM to each paper were used. These keywords are the Medical Subject Headings (MeSH Terms), a controlled vocabulary that consistently classifies each document in PubMed. First, a list of terms was collected from Category C (Diseases) in the 2017 MeSH Tree. Only keywords describing human diseases caused by DNA mutations were used to define the domains at risk of using CRISPR, including terms for cancers, infections (such as HIV), and inheritable monogenic diseases.[8] Next, all papers in PubMed from 2007-2016 containing the MeSH Terms for the at-risk diseases were identified. Papers were restricted to original scientific articles and do not include documents like reviews, news, or other non-experimental articles. From there, the database was further restricted to papers (and associated authors) that contained both a disease MeSH term and a CRISPR MeSH term.[9] The final database contains the CRISPR papers in 228 disease domains published between Q1 2013 – Q4 2016.[10] Because of the different CRISPR entry dates, some disease domains appear earlier than others for N = 442 disease-quarters in the database.

Within the 228 disease domains, there are 611 papers containing both disease and CRISPR MeSH Terms. For each author on the 611 joint papers, his or her publication history in a disease domain and CRISPR was constructed using the following procedure.[11] First for each disease domain, a sub-database at the author-paper level was constructed containing every person in PubMed that authored a paper in the focal disease or CRISPR (as defined by the MeSH Terms) from 2007 - 2016.

Second, within that sub-database, author names across papers were matched using full first names, last names, and middle initials. In order to mitigate false matches across papers, authors with last names in the top 5% of all author names had to have exact matches for the last name, first name, and middle initial in order to be considered the same person. Authors with last names in the lower 50% of all author names (very uncommon names) only had to have their last name and first

[8] Monogenic diseases, although rare, affect a wide range of cell types and are identified by a well-known inherited single mutation in a gene, for example sickle cell anemia. These are prime targets for CRISPR since there is only one gene to edit and both the mutation and the correct sequence are known.

[9] Because CRISPR was unexpected, the MeSH Terms for the tool did not appear immediately. They were added to papers in 2013, however. A check for CRISPR in the abstracts did not reveal any earlier disease domains or papers.

[10] The list of disease domains and associated MeSH terms are on file with the author.

[11] Papers are generally exclusive to one disease category.

initial match to be considered the same person. All others required an exact match of the full first name and last name to be considered the same person.[12]

Third, scientists were classified as CRISPR-disease paper authors if they authored one of the 611 papers that contained both disease and CRISPR MeSH Terms. Records from all other authors were dropped, leaving only the publication histories of the CRISPR-disease paper authors.

Fourth, using MeSH Terms, these remaining papers were classified as CRISPR Only (if the paper only had CRISPR MeSH Terms), Disease Only (if the paper only had Disease MeSH Terms), or CRISPR-disease (if the paper had CRISPR and Disease MeSH Terms). A CRISPR-disease author was classified as an external CRISPR specialist if he or she published a CRISPR Only paper first before any CRISPR-disease papers or Disease Only papers.[13]

Finally, authors with no identifiable publication history in CRISPR or the disease, usually graduate students or non-key contributors, were dropped from the sub-database. All non CRISPR-disease papers were dropped as well to focus only on the published innovations with CRISPR in each disease domain.

This process was repeated for each disease domain for a total of 228 sub-databases. These were joined into one large author-paper-disease level database with 3,019 authors, 611 joint papers, and 228 disease domains. The final database used for the analyses collapses the data to the disease-quarter level by calculating the number of total authors, external CRISPR specialists, and papers present in each quarter for each disease domain from Q1 2013 – Q4 2016 in 442 observations. The majority of disease-quarters only contain one paper, so the final database is similar to one constructed at the paper level. More than half the disease domains only have one CRISPR-disease paper over the entire time period.

5.2 Measures

The main dependent variable is a measure of external tool specialist know-how that is used to generate innovations with the tool in a domain. *Share of External CRISPR Specialists* is calculated as the number of external CRISPR specialist authors on CRISPR-disease papers divided by the

[12] The code for the grouping algorithm is on file with the author.

[13] The majority of authors in a sub-database only have one CRISPR paper in a disease domain.

total number of authors on CRISPR-disease papers. External CRISPR specialists are those that published in CRISPR first before publishing in the disease or a CRISPR-disease paper. This is measured quarterly by disease domain but due to the small number of CRISPR-disease papers, almost all disease-quarters only contain one paper. Therefore, this measure is the relative participation of external CRISPR specialists on teams that create new papers (or innovations) with CRISPR in a disease.

The main independent *Edit Difficulty* measures are binary and are different indicators of how difficult affected cells are to edit in each disease domain. The difficulty of cell editing can be measured by biological factors of the cells primarily targeted by each disease. Two key factors are (1) whether the cell can be edited *ex vivo* (edited outside a living organism and placed back in) and (2) whether the cell can self-replicate. If the cell a disease targets cannot be edited *ex vivo* or if the cell does not self-replicate, then it will be far more difficult to use CRISPR in that disease (see Section 3). The variable, *No Ex Vivo* is equal to 1 if the target cells cannot be edited *ex vivo* and 0 otherwise. The variable *No Cell Replication* is equal to 1 if the target cells do not replicate easily on their own and 0 otherwise.

To code these two variables, for each disease domain the primary target cell category was determined using information from a number of sources including the Online Mendelian Inheritance in Man database (OMIM), the Genetic and Rare Disease Information Center (GARD), and a set of gene editing review articles (LaFountaine et al. 2015; Barrangou and Doudna 2015; Cox et al. 2015; Kelton et al. 2016; Riordan et al. 2016; Scott and DeFrancesco 2016; Wang et al. 2016; Xiong et al. 2016; Bachtarzi 2017; Pandey et al. 2017; Singh et al. 2017; Song et al. 2017; Bakhrebah et al. 2018).[14] Then for each target cell category *No Ex Vivo* (National Academies of Science 2018; Cox et al. 2015) and *No Cell Replication* (Weizmann Institute of Science 2018) were coded. For example, diseases that target blood or T-cells have *No Ex Vivo* and *No Cell Replication* both equal to 0 since they are easiest to edit. Diseases that target muscle tissues have *No Ex Vivo* equal to 1 and *No Cell Replication* equal to 0 since effective editing is done within the organism. Diseases that target neurons have *No Ex Vivo* and *No Cell Replication* both equal to 1 since they are hardest to edit.

[14] The associated cell category for each disease domain and the coding for *No Ex Vivo* and *No Cell Replication* are on file with the author.

The independent variable *Quarters from First Pub* is the difference in quarters from the publication of the first CRISPR paper in a disease domain to the publication quarter of the focal CRISPR paper in the disease. The publications quarters for CRISPR-disease papers range from Q1 2013 – Q4 2016 since there are no human disease applications for CRISPR prior to 2013.

Finally, the independent variable *Total Disease Pubs* is a proxy for the attractiveness of the disease domain in terms of attention and possible funding. It is calculated as the total number of academic articles by disease domain (not including reviews, news, and other similar documents) published from Q1 2013 through Q4 2016. Articles are considered part of a disease domain if they contain MeSH Terms for those diseases.

5.3 Summary Statistics

The summary statistics for the CRISPR-disease papers and authors in the 228 disease domains by quarter is in Table 1. Table 1 presents the statistics for all diseases and further breaks down the results by Easy Diseases and Difficult Diseases as defined by whether the target cell can be edited *ex vivo* or not. On average by quarter, about 25% of the authors are external CRISPR specialists reinforcing the idea that domain-specific knowledge is important to the production of innovation in the domain but indicating that external CRISPR know-how plays a meaningful role across domains. The share of external CRISPR specialists is higher for teams writing CRISPR papers in difficult diseases at 33% versus 22% for easier diseases.

The average team size on all CRISPR-disease papers is just under 7 people. Yet the make-up of the team depends on the edit difficulty of the disease. For example, CRISPR papers in more difficult diseases have more CRISPR tool specialists but also have fewer people overall, making the individual contributions of external CRISPR specialists larger.

Further, a minority of diseases by quarter are those where CRISPR is more difficult to use. For example, 31% of diseases in a quarter have target cells that cannot be edited *ex vivo* and 17% have target cells that do not self-replicate. On average, CRISPR-disease papers are published in Q4 2015, supporting the fact that a number of diseases did not receive CRISPR until 2016, especially difficult diseases. Also, additional papers generally occur more than six months after the first CRISPR-disease paper.

[Insert Table 1 about here]

Table 2 provides the raw counts of authors and papers by the difficulty of cell editing (*No Ex Vivo*) and year of CRISPR-disease paper publication. The results suggest that there was a delay in attracting authors and publications in the more difficult diseases as compared to their easier counterparts. However, the share of external CRISPR specialists on teams publishing in difficult diseases quickly outpaced the share in easier diseases.

[Insert Table 2 about here]

This shift in more difficult diseases for authors and publications is due in part to the delayed entry of CRISPR into difficult to edit diseases. Figure 3a plots the number of diseases by the year CRISPR entered and by whether the target cell can be edited *ex vivo*. Figure 3b plots the number of diseases by the year CRISPR entered and by whether the target cell can self-replicate. For diseases where the target cells are more difficult to edit, CRISPR took additional time to enter as compared to the easier to edit diseases. Although, diseases that cannot be edited *ex vivo* overtook easier to edit diseases in 2016.

[Insert Figures 3a and 3b about here]

6 Results

6.1 Disease Edit Difficulty and Share of External CRISPR Specialists

As a way to explore where the match occurs between internal domain teams and external tool specialists for the earliest innovations, it is possible to test the relationship between disease edit difficulty and share of external tool specialists. If domain teams and external tool specialists match most often to solve *simpler problems* where the tool can be applied more quickly to a larger range of domains, then the share of external CRISPR specialists on CRISPR paper teams should *decrease* for hard to edit diseases on average. If internal domain teams and external tool specialists match most often instead to solve *complex and influential problems*, the share of external CRISPR specialists on CRISPR paper teams should *increase* for hard to edit diseases on average.

First, to establish that external tool specialists are used across domains, Figure 4 plots the distribution of the dependent variable, *Share of External CRISPR Specialists*, for publishing teams by disease edit difficulty (here where the affected cell can be edited *ex vivo*). External tool specialists are part of teams in domains that are both easier and difficult to edit, but the distribution is shifted towards a higher share of external CRISPR specialists for more difficult to edit domains.

[Insert Figure 4 about here]

Second, it is important to note that there are an increasing number of difficult to edit diseases with CRISPR availability and more available external authors with CRISPR know-how over time. Due to these facts, it should be expected that the share of external CRISPR specialists should increase over time for all diseases as the stock of CRISPR specialists increases each quarter. Figure 5a shows the increase in the number of authors over time for both easier to edit and difficult to edit diseases. Figure 5b shows similar trends for the number of external tool specialists on the research teams over time.

[Insert Figures 5a and 5b about here]

Figure 1A shows the relationship between the timing of publications and the share of external CRISPR specialists. Between Q4 2014 and Q4 2016 the share of external CRISPR specialists increases each year with some cyclicality over quarters.

Using the above facts, Table 3, Models 1-3 show fixed effect OLS regressions at the disease-quarter level to look at the effect of disease edit difficulty on the share of external CRISPR specialists contributing to new CRISPR-disease papers. From Table 3, Model 1, diseases with target cells that cannot be edited *ex vivo* have an increased share of external CRISPR specialists. The direction is different for diseases with target cells that cannot self-replicate, although the estimate is much lower and is not significant (Table 3, Model 2). Taken together the different types of edit difficulty do play different roles in affecting the percentage of external CRISPR authors (Table 3, Model 3). Here, the effect of being a *No Ex Vivo* disease is enhanced by controlling for being a *No Cell Replication* disease.

The first three models in Table 3 run OLS specifications and control for the stock of external CRISPR specialists through the quarter of publication, the number of quarters the publication is from the first CRISPR paper in a disease, and the attractiveness of the disease domain. Table 3, Models 4 – 6 have the same covariates and dependent variables but use a GLM with binomial family and logit link to account for the fractional outcome as discussed in Section 4. Although the coefficients have a different interpretation, as expected, their direction and significance are largely the same. The marginal effects of these coefficients are very similar to those of the standard linear models.

[Insert Table 3 about here]

Overall, the results support the idea that the successful match between internal domain teams and external tool specialists to create new innovations occurs in the difficult diseases where the problems are complex but solutions are highly influential. These are the diseases that had no prior viable gene therapy alternatives.

6.2 Share of External CRISPR Specialists in Subsequent Innovations

If the increase in the share of external CRISPR specialists was a simple transfer of tacit information, it might be expected that the effect would attenuate for CRISPR-disease papers after the first. Once the know-how is in the domain, it should not be necessary to have external CRISPR specialists on the team. However, contrary to this expectation, the share of external CRISPR specialists increases for subsequent innovations in more difficult diseases.

Figure 6 shows the differences in means for the share of external CRISPR specialists on CRISPR-disease papers for diseases that are more difficult to edit and those that are easier (by *ex vivo* editing), over specific periods of time. For example, the first pair of bars illustrates the difference in the average percentage of external CRISPR specialists between diseases where target cells can be edited *ex vivo* and not for the very first paper. CRISPR papers in difficult diseases (no *ex vivo* editing) have a higher share of external CRISPR specialists when only considering the first CRISPR paper published in a disease. The next pairs show the difference in means for additional CRISPR papers in a disease after the first one (i.e., excluding the first) in the first year, in the next year after, and in the remaining years. More difficult diseases have higher average shares of external

CRISPR specialists, increasing over each time period. Table A1 provides the underlying data for Figure 6.

[Insert Figure 6 about here]

Because Figure 6 highlights the results from simple T-tests, Table 4a considers the interaction between subsequent innovations after the first and the edit difficulty of the disease while controlling for disease attractiveness as well as quarter, age, and disease fixed effects. The results in Table 4a echo the trends found in Figure 6. Table 4a, Model 1 includes an interaction term between the time distance of an additional innovation in a disease and disease edit difficulty (using the *No Ex Vivo* measure). The results suggest that for diseases that are more difficult to edit and as additional innovation is more time-distant from the original joint paper, the share of authors with external CRISPR specialization increases.

Table 4a, Model 2 breaks these interactions down further by the length of time the innovation appeared after the first. As more time passes after the first paper, the share of external CRISPR specialists continues to increase during this very early stage. Whether this pattern will continue as the tool matures is the subject of a future study as more time is allowed to pass. Table 4a, Models 3-4 show results in the same direction and significance levels for the analogous GLMs. Figure 7 plots the coefficients in Table 4a, Model 2 to illustrate the increasing trend in the share of external CRISPR specialists for subsequent innovations.

[Insert Table 4a and Figure 7 about here]

Because some diseases only have one CRISPR paper, it might be a concern that these diseases are different and influencing the results in Table 4a. In order to account for that, Table 4b performs an identical analysis but only for the 93 diseases that have more than one CRISPR paper in the dataset. The results are very similar, so including the entire dataset is not influencing the core findings for the OLS or GLM specifications.

[Insert Table 4b about here]

6.3 Future Supplementary Analyses

The results in Sections 6.1 and 6.2 are the first steps towards empirically understanding how the earliest teams acquire complementary tool know-how across domains with different levels of ex ante difficulty. However, the analysis presented suggests several other areas of inquiry that can help to further this understanding. For example, are there characteristics of external CRISPR specialists that attract some to difficult problems and others to easier ones? As more data becomes available, it will be possible to exploit heterogeneity in adopters that does not currently exist in the data to test whether the tenure status of external CRISPR specialists plays a role in their attraction to more complex and influential problems. It could be the case that tenured external CRISPR specialists have more incentives to aim for the Nobel Prize while non-tenured external CRISPR specialists just need to publish. This could cause tenured external CRISPR specialists to seek out and match in more difficult domains.

Other analyses are aimed at assessing the benefits to building versus acquiring the complementary tool know-how. Specifically, the analyses address the questions do teams that use external CRISPR specialists get to publication faster? Does this change given domain difficulty? For this, team members on CRISPR-disease papers are being matched to their CRISPR order histories at Addgene to determine the earliest date the team received CRISPR and the length of time between the first order and the eventual publication.

7 Discussion and Conclusion

When a new tool is introduced that requires some investment in tool-specific know-how, how to combine that know-how with domain-specific knowledge is a decision firms, mangers, and individual innovators face as they innovate. In general, innovators in a domain can either learn to use the tool themselves or can acquire the complementary tool know-how from an external source. For example, firms face this choice when deciding how to use AI and individual academics may make similar decisions when deciding how to use STATA or Python in their work. However, when external tool specialists are scarce, they also have a choice over which teams to join. They could collaborate with

teams in easy domains to expand the use of the tool as far as possible or in difficult domains where the problems are more complex but the solutions are highly influential.

Previous research has shown that access to research tools increases innovation but often access is conflated with the ability to use the tool, limiting investigations into how innovators use tools to generate innovations, especially if they are trying to be first. One mechanism this paper highlights is the role of external tool specialists in generating early innovation by providing complementary know-how needed to use the tool.

Using the introduction of the new breakthrough DNA-editing tool, CRISPR, and applying the unanticipated timing of CRISPR entry in different human disease domains, it is possible to separately identify the knowledge bases of the academic scientists responsible for the new articles that use CRISPR in a disease. Because the analysis focuses on the first days of CRISPR (2012-2016), these papers represent the earliest innovations in each domain with the tool. It is also possible to create measures for the difficulty of applying CRISPR in the different disease domains. Human diseases primarily target certain cell categories and each category is easier or harder to edit based on biological factors specific to the cells. For example, CRISPR is more difficult to apply in target cells that cannot be edited *ex vivo* or cannot self-replicate.

The variation in the difficulty of applying CRISPR provides a novel lever to presents a direct mechanism for how new tools are adopted and incorporated into early innovations. First, a higher share of external CRISPR tool specialists participate in early innovations with the tool in difficult disease domains. This suggests the match between internal domain teams and external tool specialists occurs more often in domains with complex and influential problems. To understand this result, consider the different experiences of CRISPR adoption in HIV and muscular dystrophy. The human immunodeficiency virus (HIV) had one of the earliest introductions of CRISPR in part due to the targeted T cells being easy to edit and the large amount of previous research conducted on gene editing alternatives. The first study using CRISPR to make new advances in HIV was published in 2013 by Yoshio Koyanagi, the PI of a Viral Pathogenesis lab at Kyoto University in Japan (Ebina et al. 2013). He and his three co-authors, all part of the lab, had previous experience in HIV, but none specifically in CRISPR. Although external CRISPR know-how would be useful in this case, the internal team ordered CRISPR from Addgene and modified it for their application but did not collaborate on the paper with an external CRISPR tool specialist outside of the domain. In

contrast, muscular dystrophy targets muscle cells that are more challenging to edit and success would represent an enormous advance in medicine. For the first paper that successfully used CRISPR inside a mouse to treat Duchenne muscular dystrophy, Charles Gersbach, a leading muscular dystrophy researcher at Duke University and his lab were having problems delivering CRISPR to the nucleus of the target cells. To overcome this problem, the team incorporated the knowledge of Feng Zhang, a CRISPR co-founder, to create a new delivery solution. The resulting paper lists both professors as contributing authors (Nelson et al. 2016; Duke Today Staff 2015). The example suggests that authors in internal teams working on difficult and influential diseases are more likely to look for and attract external CRISPR specialists to collaborate.

Second, there is evidence that the higher share of external CRISPR specialists persists for subsequent innovations in more difficult disease domains. This result is not immediately intuitive. If research in the domain using CRISPR became easier after the tool was first introduced, then complementary know-how about the tool should be less valuable and external tool specialists less necessary. However, in this setting, the ultimate goal is to use CRISPR to create commercial therapies and drugs for human use. In order to do this, additional research in each disease will try to use the tool in increasingly complex organisms. Even within mammals, as the organism gets closer to humans, editing becomes more difficult and the solutions more notable, attracting a higher share of external CRISPR specialists. As an example, one of the first CRISPR experiments in muscular dystrophy was to deliver CRISPR inside a living mouse (Nelson et al. 2016). The next step was to deliver CRISPR inside living dogs (Amoasil et al. 2018). Note that the author teams for each organism are different. This pattern of research is repeated in other diseases with more difficult to edit target cells.

Although the current study was conducted in the context of CRISPR, the overarching findings have implications for firms and individuals thinking about when and how to adopt new tools for innovation. Variations in the ex ante domain difficulty and solution novelty is not unique to CRISPR and research teams looking to be early adopters in tools like AI will have to weigh the complexity and influence of their goals when considering how best to attract and collaborate with external tool specialists. In order to be first to innovate with a new tool, external specialists are not always necessary to acquire complementary tool know-how.

However, external tool specialists may be more likely to find successful matches with internal domain teams that focus on more complex problems with highly influential solutions.

The findings contribute to the literature on innovative teams and team structure by uniquely showing that not just features of management, organizational structure, or industry are important for effective team design. Team composition is also driven by the specific nature of the problem to be solved and the nature of tools available for innovation. This paper is one of the first in a series that uses CRISPR to answer key questions in innovation, management, and economics. For example, future papers can build on the results established here to study the effect of breakthrough technologies on academic entrepreneurship, the impact of policies regarding genetically modified organisms on agricultural product development, the effect of unresolved intellectual property disputes on scientific innovation, and how to incentivize ethical innovation without stifling important technological advances.

8 References

Adams, J. D., G. C. Black, J. R. Clemmons, and P. E. Stephan. 2005. "Scientific Teams and Institutional Collaborations: Evidence from US Universities, 1981–1999." Research Policy 34 (3): 259–85.

Ahuja, G. and R. Katila. 2001. Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study. Strategic Management Journal, 22(3): 197-220.

Amoasi, L., J. Hildyard, H. Li, E. Sanchez-Ortiz, A. Mireault, D. Caballero, R. Harron, T. Stathopoulou, C. Massey, J. Shelton, R. Bassel-Duby, R. Piercy, and E. Olson. 2018. "Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy." Science, 362(6410):86-91.

Arora, A., and A. Gambardella. (1994). The changing technology of technological change: general and abstract knowledge and the division of innovative labour. Research Policy, 23(5), 523-532.

Azoulay, P., W. Ding, and T. Stuart. 2009. "The Effect of Academic Patenting on the Rate, Quality, and Direction of (Public) Research Output." *Journal of Industrial Economics* 57(4): 637-676.

Bachtarzi, H. 2017. "Ex vivo and in vivo genome editing: a regulatory scientific framework from early development to clinical implementation." Regenerative Medicine, 12(8): 1015-1030.

Bakhrebah, M.A., M.S. Nassar, M.S. Alsuabeyl, W.A. Zaher, and S.A. Meo. 2018. "CRISPR technology: new paradigm to target the infectious disease pathogens." European Review for Medical and Pharmacological Sciences, 22: 3448-3452.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero D, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science. 315(5819): 1709–1712.

Barrangou, R. and J. Doudna. 2016. "Applications of CRISPR technologies in research and beyond." Nature Biotechnology, 34(9): 933-941.

Bikard, M., Murray, F., Gans, J. (2015). Exploring Tradeoffs in the Organization of Scientific Work: Collaboration and Scientific Reward. Management Science, 61(7): 1473 -1495.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-Type III effectors. Science. 326(5959):1509-1512.

Bresnahan, T. F., and M. Trajtenberg. 1995. "General Purpose Technologies: Engines of Growth?" Journal of Econometrics 65(1): 83-108.

Cassiman, B. and R. Veugelers (2006) "In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition." Management Science. 52: 68-82.

Coase, R.A. (1937). "The nature of the firm". Economica 16 (4), 386–405.

Cockburn, I.A., R. Henderson, and S. Stern. 2017. "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis." Ajay K. Agrawal, Joshua Gans, and Avi Goldfarb, eds. *The Economics of Artificial Intelligence: An Agenda.* Chicago University Press: Chicago, IL. Forthcoming.

Cohen, W. M. and D.A. Levinthal (1989). Innovation and learning: the two faces of R & D. The Economic Journal, 99(397), 569-596.

Cohen, W. M. and D.A. Levinthal (1990). "Absorptive Capacity: A New Perspective on Learning and Innovation." Administrative Science Quarterly. 128-152.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science. 339(6121): 819-823.

Cox, D., R. Platt, and F. Zhang. 2015. "Therapeutic Genome Editing: Prospects and Challenges," Nat Med. 2015 Feb; 21(2): 121–131.

Cummings, J. N., & Kiesler, S. (2007). Coordination costs and project outcomes in multiuniversity collaborations. Research Policy, 36(10), 1620-163.

Cyranoski, D. 2018. "CRISPR-baby scientist fails to satisfy critics." Nature, 564, 13-14.

Dallon, A. 2017. Everything you need to know about Waymo's self-driving car project. Digital Trends (April 26). Available at: https://waymo.com/journey/.

David, P. 1990. "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox." American Economic Review 80(2): 355-36.

Doudna J (2015) Genomic engineering and the future of medicine. J. Amer. Medical Assoc. 313(8):791-792.

Doudna, J. and S. Sternberg. 2017. *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution*. Houghton Mifflin Harcourt. New York: New York.

Duke Today Staff. 2015. "CRISPR Treats Genetic Disorder in Adult Mammal." Duke Today, December 31. Available at: https://today.duke.edu/2015/12/crisprmousedmd

Ebina, H., N. Misawa, Y. Kanemura, and Y. Koyanagi. 2013. "Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus." Scientific Reports, 3: 2510.

Fan M, Tsai J, Chen B, Fan K, LaBaer J (2005) A central repository for published plasmids. Science. 307(5717): 1877.

Fleming, L. (2001) "Recombinant Uncertainty in Technological Search." Management Science, 47: 117-132.

Furman J, and S Stern (2011) Climbing atop the shoulders of giants: The impact of institutions on cumulative knowledge production. Amer. Econom. Rev. 101(5): 1933-1963.

Furman, J and F. Teodoridis (2018) "The Cost of Research Tools and the Direction of Innovation: Evidence from Computer Science and Electrical Engineering." Working.

Gale, D. and L. Shapley. 1962. "College Admissions and the Stability of Marriage," American Mathematical Monthly, 69, 9-15.

Gans, J. and F. Murray. 2014. Markets for Scientific Attribution. NBER Working Paper 20677. Available at: http://www.nber.org/papers/w20677

Gantz V, Jasinskiene N, Tatarenkova O, Fazekas A, Macias V, Bier E, James A (2015) Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. PNAS. 112(49): E6736–E6743.

GARD. Genetic and Rare Diseases Information Center. Available at: https://rarediseases.info.nih.gov/

Grigoriou, K. and F. Rothaermel. 2017. "Organizing for Knowledge Generation: Internal Knowledge Networks and the Contingent Effect of External Knowledge Sourcing." Strategic Management Journal. 38: 395-414.

Griliches, Z. (1957) Hybrid corn: An exploration in the economics of technological change. Econometrica. 25(4):501-522.

Griliches, Z. (1958) Research costs and social returns: Hybrid corn and related innovations. J. Political Econom. 66(5): 419-431.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 337(6096): 816-21.

Jones, B. F. 2009. "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *Review of Economic Studies* 76(1): 283-31

Kaiser, J. 2016. "The Gene Editor CRISPR won't fully fix sick people anytime soon. Here's why." Science. May. 3, 2016

Kaplan, S. and K. Vakili (2015) "The Double-edged Sword of Recombination in Breakthrough Innovation." Strategic Management Journal. 36: 1435-1457.

Kelton, W., T. Pesch, S. Matile, and S. Reddy. 2016. "Surveying the Delivery Methods of CRISPR/Cas9 for ex vivo Mammalian Cell Engineering." CHIMIA, 70(6): 439-442.

LaFountaine, J., K. Fathe, H. Smyth. 2015. "Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9." International Journal of Pharmaceutics. 494: 180–194

Ledford H (2015) CRISPR: The disruptor. Nature. 522(7554): 20-24.

Ma H, Marti-Gutierrez N, Park SW, Wu J, Lee Y, Suzuki K, Koski A, Ji D, Hayama T, Ahmed R, Darby H, Dyken C, Li Y, Kang E, Park AR, Kim D, Kim ST, Gong J, Gu Y, Xu X, Battaglia D, Krieg S, Lee D, Wu D, Wolf D, Heitner S, Belmonte JC, Amato P, Kim JS, Kaul S, Mitalipov S. (2017) Correction of a pathogenic gene mutation in human embryos. Nature. 548(7668): 413-419.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. (2013) RNA-guided human genome engineering via Cas9. Science. 339(6121): 823-6.

Marr, B. 2016. The Top 10 AI And Machine Learning Use Cases Everyone Should Know About. Forbes (Sep 30).

Method (2012) Method of the year 2011. Nat. Methods. 9: 1.

MeSH Terms. 2017. PubMed Database, MeSH Tree. Available at: ftp://nlmpubs.nlm.nih.gov/online/mesh/2017/ .

Mokyr, J. 2002. *Gifts of Athena: Historical Origins of the Knowledge Economy.* Princeton University Press.

Moscou M, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science. 326(5959): 1501.

Moser, P. 2005. "How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World Fairs." American Economic Review 95(4): 1214-1236.

Mowery, D.C., A.A. Ziedonis. 2007. Academic patents and materials transfer agreements: substitutes or complements? J. Technol. Transfer 32 (3), 157–172.

Murray, F. 2002. "Innovation as co-evolution of scientific and technological networks: exploring tissue engineering." Research Policy, 31(8-9): 1389-1403.

Murray, F., Aghion P, Dewatripont M, Kolev J, Stern S (2016) Of mice and academics: Examining the effect of openness on innovation. Amer. Econom. J.: Econom. Policy. 8(1): 212-252

Nagle, F. and F. Teodoridis. 2017. "Jack of All Trades and Master of Knowledge: The Role of Generalists in Novel Knowledge Integration." USC Dornsife Institute for New Economics Thinking, Working Paper No. 17-23.

National Academies of Science. 2018. *Human Genome Editing: Science, Ethics, and Governance.* Available at: https://www.nap.edu/read/24623/chapter/6#84

Nelson, C., C. Hakim, D. Ousterout, P. Thakore, E. Moreb, R. Castellanos Rivera, S. Madhavan, X. Pan, F. Ran, W. Yan, A. Asokan, F. Zhang, D. Duan, C. Gersbach. 2016. "In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy." Science, 351(6271):403-7.

Nelson, R.R. (1981). "Research on productivity growth and productivity differences: Dead ends and new departures". Journal of Economic Literature 19 (3), 1029–1064.

Nelson, R.R. (2003). "On the uneven evolution of human know-how". Research Policy 32 (6), 909–922.

OMIM. Online Mendelian Inheritance in Man Database. Available at: https://www.omim.org/

Pandey, V., A. Tripathi, R. Bhushan, A. Ali, and P. Dubey. 2017. "Application of CRISPR/Cas9 Genome Editing in Genetic Disorders: A Systematic Review Up to Date." Journal of Genetic Syndromes and Gene Therapy, 8(2): 321-331.

Papke, L.E., Wooldridge, J.M., 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. Journal of Applied Econometrics 11, 619–632.

Pennisi E (2013) The CRISPR craze. Science. 341(6148): 833-836.

Pinzon, E. (ed.) 2015. *Thirty Years with Stata: A Retrospective.* College Station, TX: Stata Press

Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F (2014) CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell. 159(2):440-55.

Polanyi, M. 1962. "Tacit knowing: Its bearing on some problems of philosophy." Reviews of Modern Physics, 34(4): 601.

PubMed. 2017-2018. Available at: https://www.ncbi.nlm.nih.gov/pubmed.

Rabinow, P. 2011. *Making PCR: A Story of Biotechnology.* Chicago, IL: University of Chicago Press.

Regalado A (2014) Who owns the biggest biotech discovery of the century? MIT Technology Review (December 4) http://www.technologyreview.com/featuredstory/532796/who-owns-the-biggest-biotech-discovery-of-the-century/.

Regalado, A. 2016. "Can CRISPR Save Ben Dupree?", MIT Tech Review, Vol. 119. No. 8 pp. 81-87.

Riordan, S., D. Heruth, L.Q. Zhang, and S.Q. Ye. 2015. "Application of CRISPR/Cas9 for biomedical discoveries." Cell and Bioscience, 5:33-44.

Rogers E (1995[1962]) Diffusion of Innovations. 4 ed. (The Free Press: Simon & Schuster, New York, NY).

Rosenberg, N. (1982). Inside the Black Box: Technology and Economics. Cambridge University Press, Cambridge.

Rosenberg, N. (1994). Exploring the Black Box: Technology, Economics, and History. Cambridge University Press, Cambridge.

Rosenberg, N. (2009). "Some critical episodes in the progress of medical innovation: An Anglo-American perspective". Research Policy 3 (2), 234–242.

Rosenberg, N., and M. Trajtenberg. 2004. "A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States." Journal of Economic History 64(1): 61-99.

Roth, A. E. 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," Journal of Political Economy, 92:991-1016.

Schilling, M. A., & Green, E. (2011). Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. Research Policy, 40(10), 1321-1331.

Scotchmer, S. 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives* 5(1): 29-41

Scott, C. T. and L. DeFrancesco. 2016. "Gene therapy's out-of-body experience." Nature Biotechnology, 34(6): 600-607.

Singh, V., N. Gohil, R.R. Garcia, D. Braddick, and C. K. Fofie. 2018. "Recent advances in CRISPR-Cas9 genome editing technology for biological and biomedical investigations." Journal of Cellular Biochemistry, 119: 81-94.

Song, M. 2017. "The CRISPR/Cas9 System: Their Delivery, In Vivo and Ex Vivo Applications and Clinical Development by Startups." Biotechnology Progress, 33(4): 1035-1045.

Specter, M. (2015) The gene hackers: A powerful new technology enables us to manipulate our DNA more easily than ever before. The New Yorker (16 November).

Stephan, P. 2012. *How Economics Shapes Science*. Cambridge, Massachusetts: Harvard University Press.

Stockton, N. 2017. "CRISPR kills HIV and eats Zia 'like Pac-man'. It's Next Target? Cancer." WIRED.

Strandberg, K. 2010. "Norms and the Sharing of Research Materials and Tacit Knowledge." *Working Within the Boundaries of Intellectual Property: Innovation Policy For The Knowledge Society.* R. C. Dreyfuss, D. L. Zimmerman, and H. First (eds.), Oxford University Press.

Teece, D. J. (1986). "Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy." Research Policy, 15(6): 285-305.

Teodoridis, F. (2017) Generalists, specialists, and the direction of inventive activity. Forthcoming, Management Sci.

Thompson, N. and S. Zyontz. (2017) Who tries (and who succeeds) in staying at the forefront of science: Evidence from the DNA-editing technology, CRISPR. Working paper.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. Science, 342(6157), 468-472

Walsh, J., A. Arora, and W. Cohen (2003) "Effects of Research Tool Patents and Licensing on Biomedical Innovation." *Patents in the knowledge-based economy.* W.M. Cohen and S. A. Merrill (eds.), Washington, DC : National Academies Press.

Walsh JP, Cohen WM, Cho C (2007) Where excludability matters: Material versus intellectual property in academic biomedical research. Res. Policy. 36(8):1184–1203.

Wang, L., F. Li, L. Dang, C. Liang, C. Wang, B. He, J. Liu, D. Li, X. Wu, X. Xu, A. Lu, and G. Zhang. 2016. "In Vivo Delivery Systems for Therapeutic Genome Editing." International Journal of Molecular Science. 17:626-645.

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu JL (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat. Biotechnol. 32(9): 947–951.

Weitzman, M. 1996. Hybridizing Growth Theory. American Economic Review, 86: 207-212.

Weitzman, M. 1998. Recombinant Growth. Quarterly Journal of Economics, 113(2):331-360.

Weizmann Institute of Science. 2018. *Bionumbers*. Available at: http://bionumbers.hms.harvard.edu/search.aspx AND http://book.bionumbers.org/how-quickly-do-different-cells-in-the-body-replace-themselves/

Whitaker, B. 2018. "CRISPR: The Gene Editing Tool Revolutionizing Biomedical Research." *60 Minutes*. Available at: https://www.cbsnews.com/news/crispr-the-gene-editing-tool-revolutionizing-biomedical-research/

Williams, H. 2013. "Intellectual Property Rights and Innovation: Evidence from the Human Genome," Journal of Political Economy 121 (1), 1-27.

Williams, H. and B. Sampat. 2015. "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome" working paper.

Williamson, O. (1975). Markets and Hierarchies. Free Press, New York, NY.

Williamson, O. (1985). The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting. Free Press, New York and London.

Wills, M. (2018) "The Evolution of the Microscope," *JSTOR Daily*, March 27. Available at: https://daily.jstor.org/the-evolution-of-the-microscope/

Wuchty, S., B. Jones, and B. Uzzi (2007) "The Increasing Dominance of Teams in Production of Knowledge." Science. 316(5827):1036-9.

Xiong, X., M. Chen, W.A. Lim, D. Zhao, and L.S. Qi. 2016. "CRISPR/Cas9 for Human Genome Engineering and Disease Research." Annual Review of Genomics and Human Genetics, 17: 131-54.

Zahra, S.A. and G. George. 2002. Absorptive capacity: A review, reconceptualization, and extension. Academy of Management Review, 27(2): 185-203.

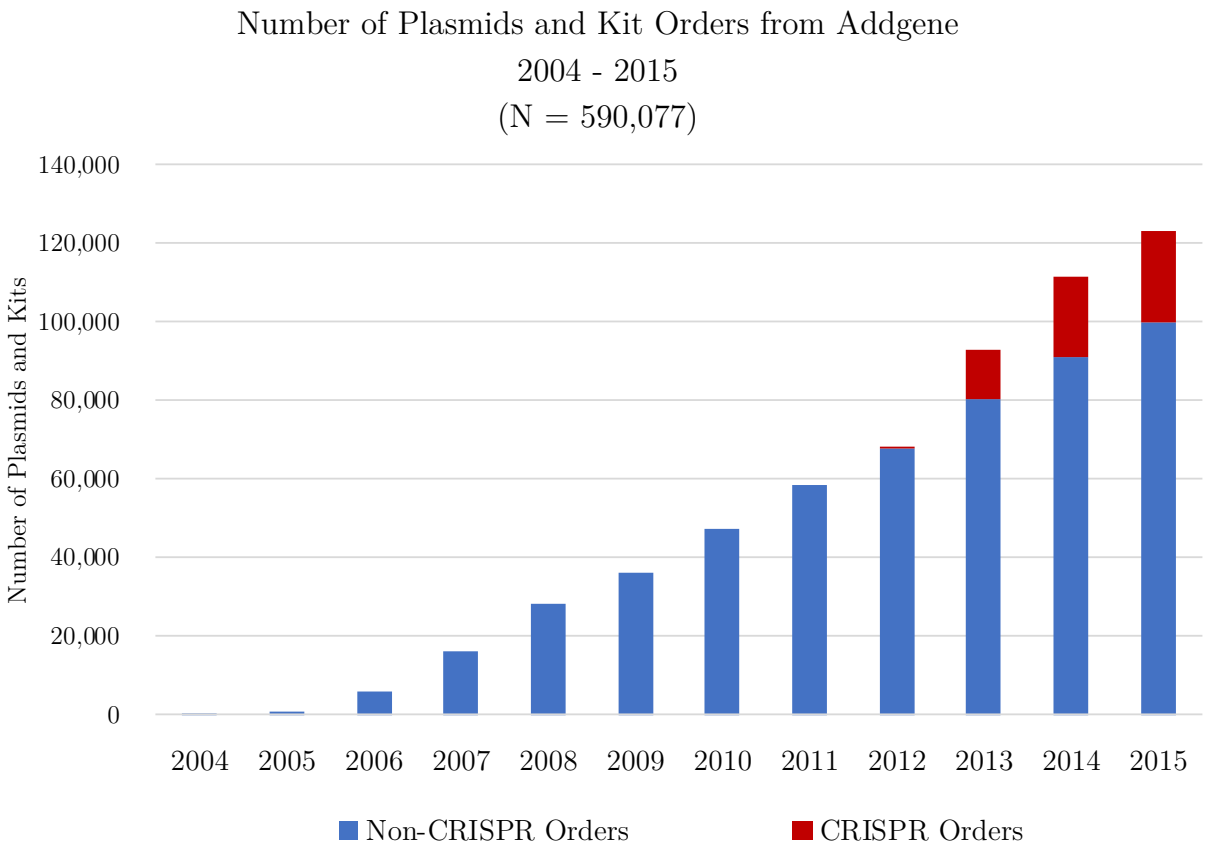
Zhang, S. 2018. "The Little-Known Nonprofit Behind the CRISPR Boom." The Atlantic (June 13). Available at: https://www.theatlantic.com/science/archive/2018/06/the-dna-depot-that-has-sent-crispr-to-62-countries/562370/.

Zucker, L. G. and M.R. Darby. 2001. "Capturing technological opportunity via Japan's star scientists: Evidence from Japanese firms' biotech patents and products." The Journal of Technology Transfer, 26(1-2): 37-58.

Zyontz, S. (2016) Technological breakthroughs, entry, and the direct of scientific progress: Evidence from CRISPR/Cas9. Working paper.

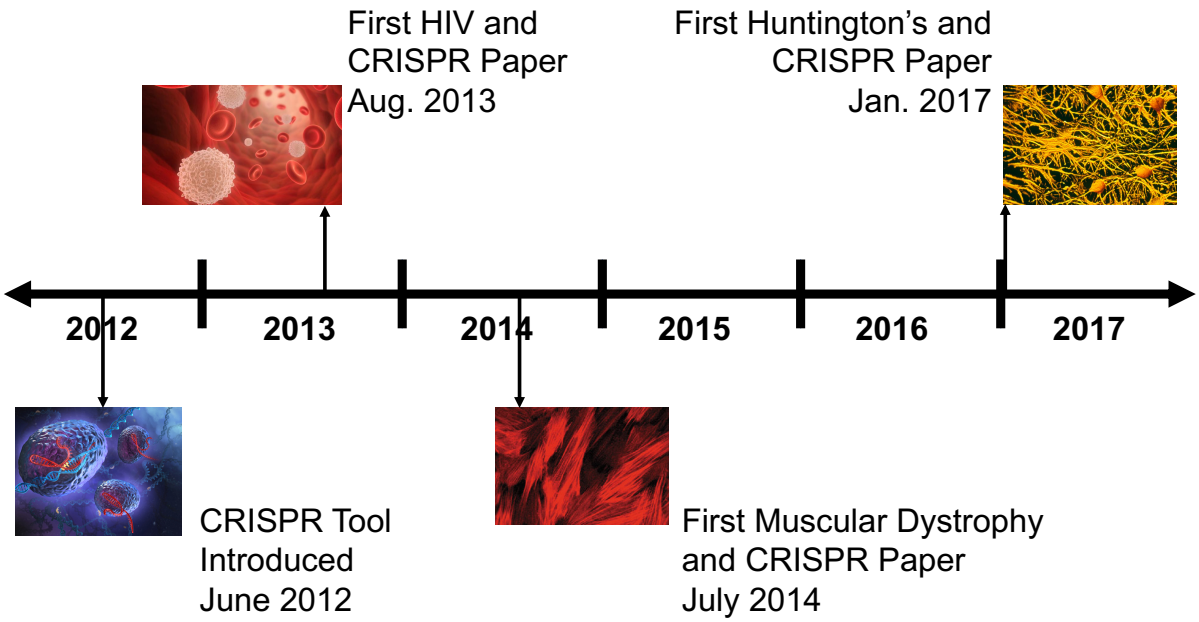
9 Figures and Tables

Figure 1. Addgene Plasmid and Kit Orders by Year

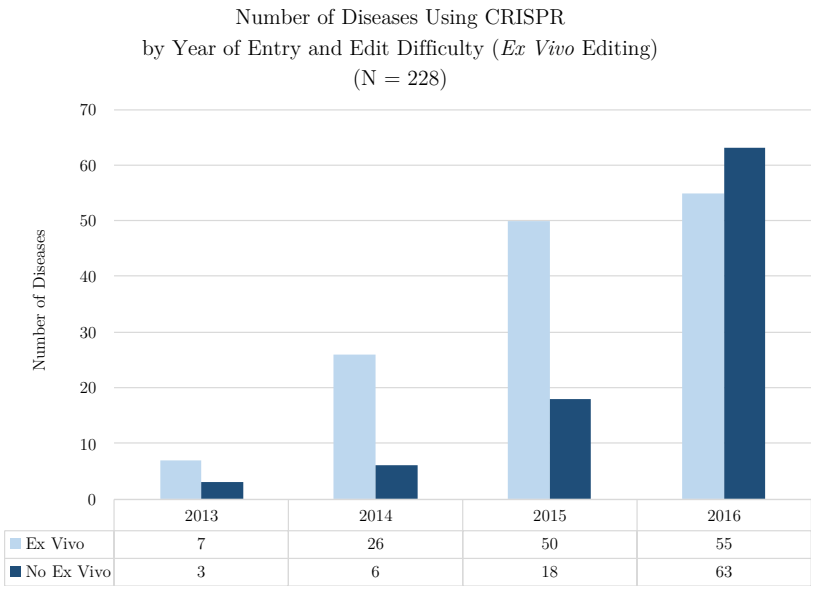


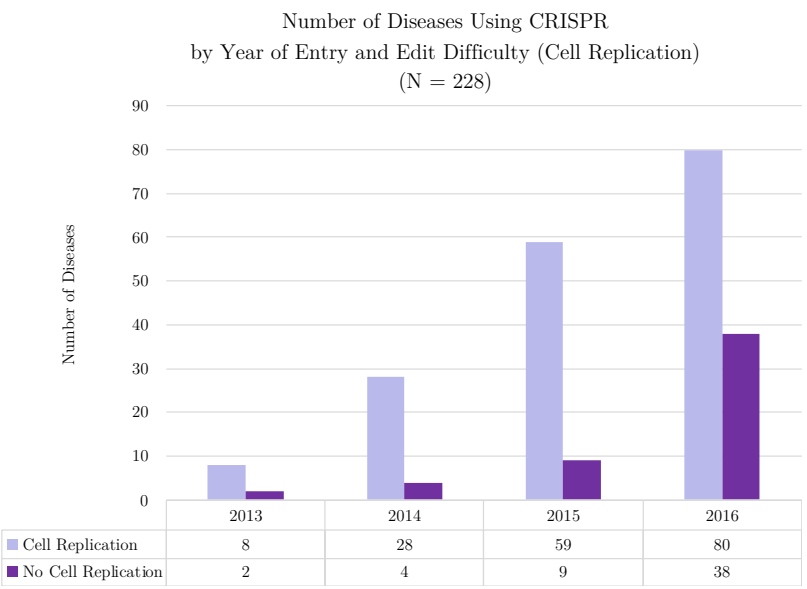
Notes. This graph shows the number of individual plasmids and sets of plasmids (kits) that Addgene sold per year from its start in 2004 through 2015. The blue bars are orders for non-CRISPR plasmids and the red bars are orders for CRISPR plasmids. Source: Addgene internal records, 2004 – 2015.

Figure 2. Example Timeline of CRISPR Introduction and Applications



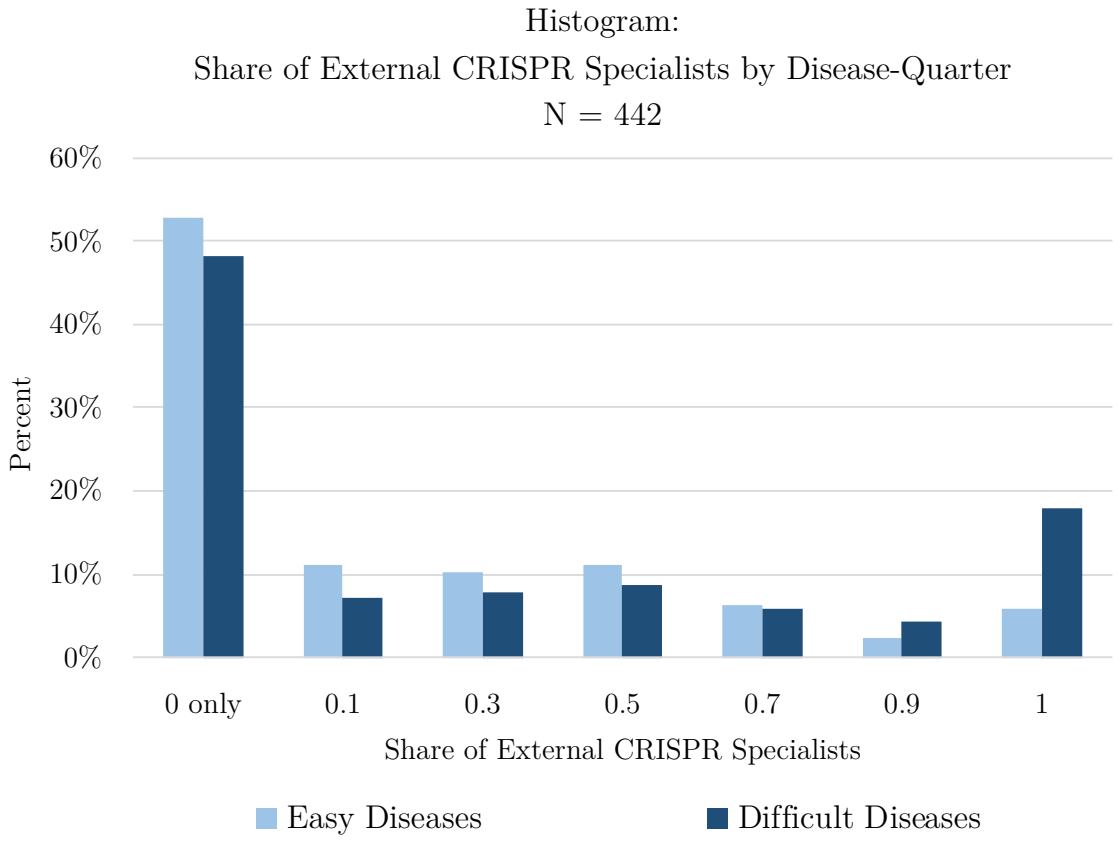
Figures 3a and b. CRISPR Entry by Year and Disease Cell Edit Difficulty





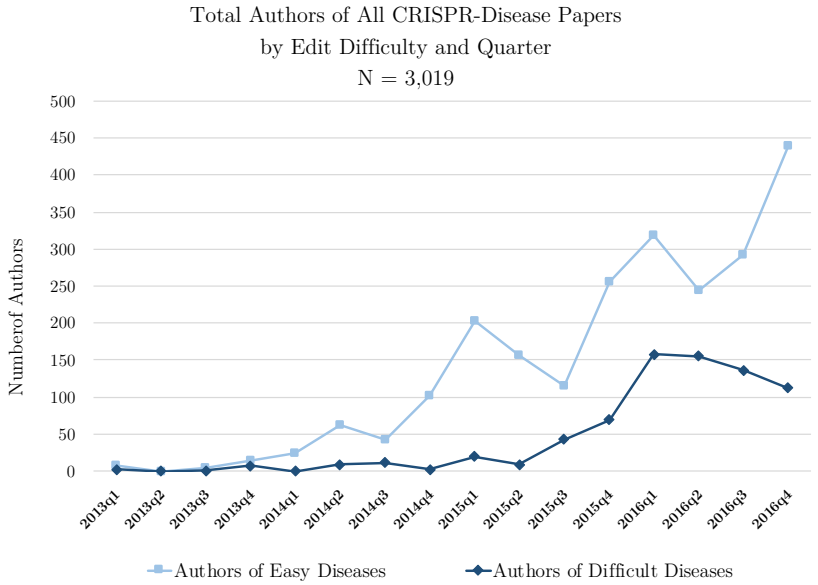
Notes. Figure 3a shows the number of diseases that first receive CRISPR each year by whether the target cell can be edited *ex vivo* or not. If the cell a disease targets cannot be edited *ex vivo*, then it will be harder and costlier (in resources) to edit. The light blue bars are the number of diseases that receive CRISPR if their affected cells can be edited *ex vivo*. The dark blue bars are the number of diseases that receive CRISPR if their affected cells cannot be edited *ex vivo*. Figure 3b shows the number of diseases that first receive CRISPR each year by whether the cell can self-replicate or not. If the cell a disease targets does not self-replicate, then it will be harder and costlier (in resources) to edit. The light purple bars are the number of diseases that receive CRISPR if their target cells do self-replicate. The dark purple bars are the number of diseases that receive CRISPR if their target cells cannot self-replicate. For both measures, CRISPR entry is delayed in the difficult to edit diseases. Source: PubMed publications and MeSH Terms, 2013 – 2016.

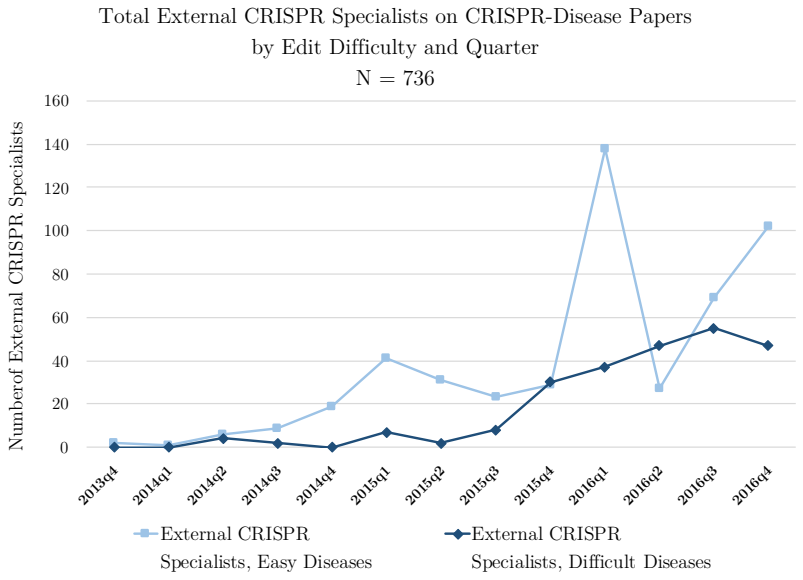
Figure 4. Distribution of the Share of External CRISPR Specialists by Disease-Quarter and Disease Cell Edit Difficulty (*Ex Vivo* Editing)



Notes. This figure shows the distributions of the share of external CRISPR specialists on teams that publish CRISPR papers in easy and difficult diseases by disease and quarter. The light blue bars represent the distribution pattern for easy diseases and the dark blue bars represent the distribution pattern for difficult diseases. The distribution of the dark blue bars shifts to the right suggesting that teams publishing in difficult diseases have a higher share of external CRISPR specialists. The difficulty of the disease is measured by whether the targeted cell can be edited *ex vivo*. If the cell a disease targets cannot be edited *ex vivo*, then it will be harder and costlier (in resources) to edit. CRISPR-Disease papers are defined as articles containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR paper in a disease is considered an external CRISPR specialist if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.

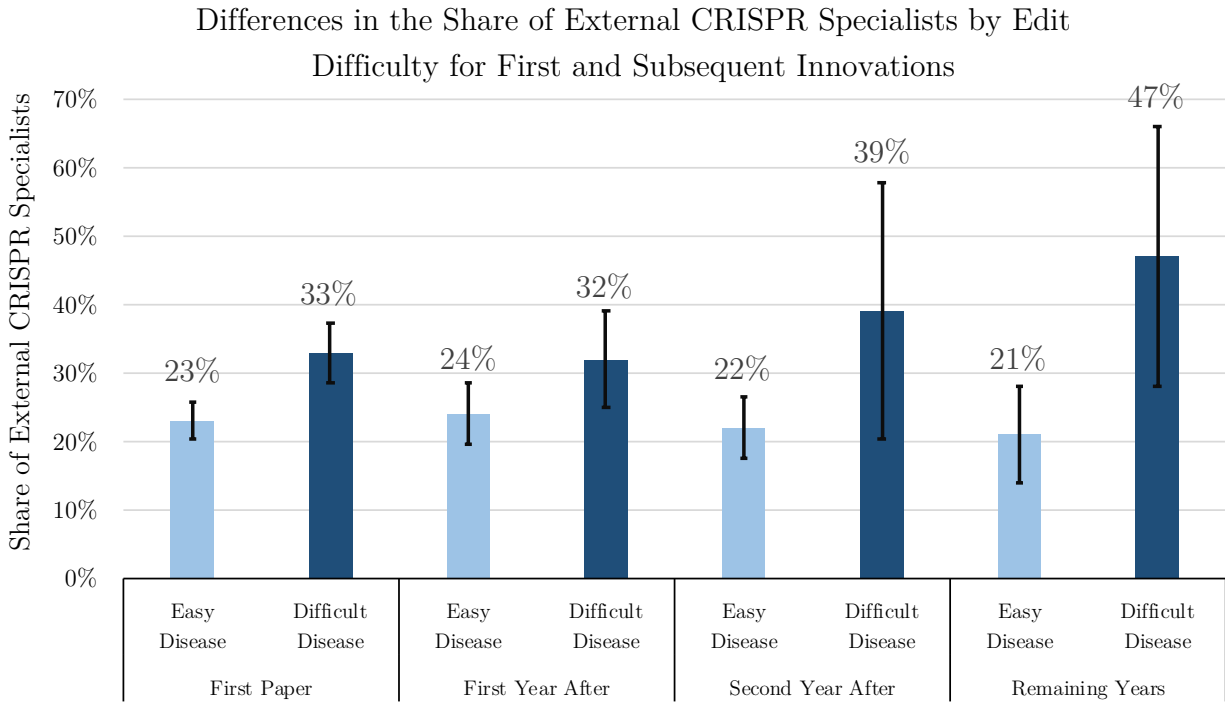
Figures 5a and b. Number of Authors on CRISPR-Disease Papers by Quarter and Disease Affected Cell Edit Difficulty (*Ex Vivo* Editing)





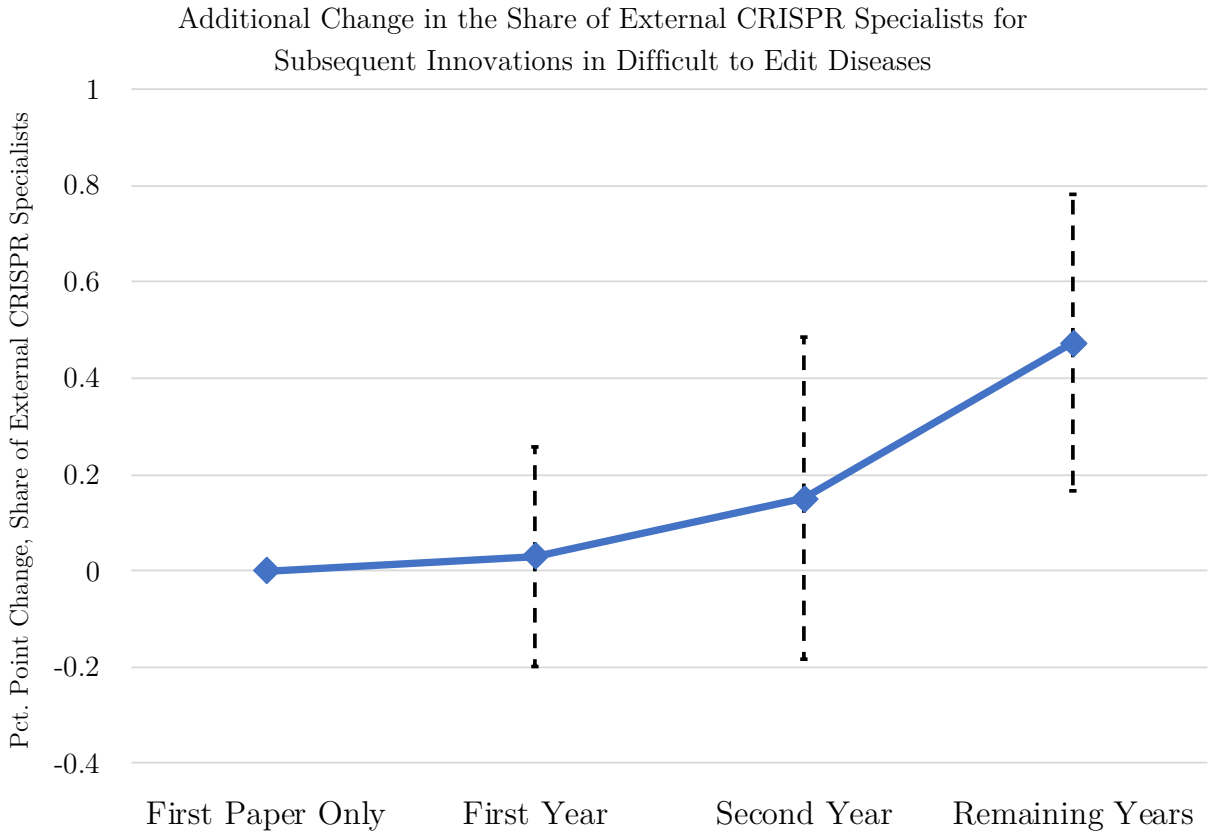
Notes. Figure 4a shows the increase in the total number of authors on CRISPR-disease papers by disease difficulty and quarter. Figure 5b shows the increase in only external CRISPR specialist authors on CRISPR-disease papers by disease difficulty and quarter. In both figures, the light blue line represents the trend for easy diseases and the dark blue line represents the trend for difficult diseases. The difficulty of the disease is measured by whether the targeted cell can be edited *ex vivo*. If the cell a disease targets cannot be edited *ex vivo*, then it will be harder and costlier (in resources) to edit. CRISPR-Disease papers are defined as articles containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR paper in a disease is considered an external CRISPR specialist if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.

Figure 6. Mean Differences in the Share of External CRISPR Specialists by Disease Cell Edit Difficulty (*Ex Vivo* Editing) for First and Subsequent Innovations



Notes. This figure shows the difference in means between the share of external CRISPR specialists for CRISPR papers in easy diseases versus difficult diseases based on whether the paper was the first or subsequent CRISPR paper in the disease. The light blue bars represent the means for easy diseases and the dark blue bars represent the means for difficult diseases. The share of external CRISPR specialists is higher on average for teams publishing CRISPR papers in difficult diseases for both the first and subsequent papers. The difficulty of the disease is measured by whether the targeted cell can be edited *ex vivo*. If the cell a disease targets cannot be edited *ex vivo*, then it will be harder and costlier (in resources) to edit. CRISPR-Disease papers are defined as articles containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR paper in a disease is considered an external CRISPR specialist if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.

Figure 7. Additional Change in the Share of External CRISPR Specialists for Subsequent Innovations in Difficult to Edit Diseases (*Ex Vivo* Editing)



Notes. This figure shows the change in the share of external CRISPR specialists on teams publishing subsequent CRISPR-disease papers in difficult diseases an corresponds to the coefficients estimated in Table 4a, Model 2. As later subsequent CRISPR papers are published in difficult diseases as compared to easy diseases, the share of external CRISPR specialists authors increases. The difficulty of the disease is measured by whether the targeted cell can be edited *ex vivo*. If the cell a disease targets cannot be edited *ex vivo*, then it will be harder and costlier (in resources) to edit. CRISPR-Disease papers are defined as articles containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR paper in a disease is considered an external CRISPR specialist if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.

Table 1. Summary Statistics

Variable	Description	All Diseases					Easy Diseases (No *Ex Vivo* = 0)					Difficult Diseases (No *Ex Vivo* = 1)				
		N	Mean	Std. Dev.	Min	Max	N	Mean	Std. Dev.	Min	Max	N	Mean	Std. Dev.	Min	Max
Share of External CRISPR Specialists	Share of authors with a CRISPR publication before a Disease or CRISPR-Disease publication	442	0.251	0.342	0	1	303	0.215	0.307	0	1	139	0.328	0.398	0	1
Total External CRISPR Specialists	Total authors with a CRISPR publication before a Disease or CRISPR-Disease publication	442	1.665	2.920	0	22	303	1.640	3.062	0	22	139	1.719	2.593	0	12
Total Authors	Total authors on a CRISPR-Disease publication	442	6.830	7.186	1	60	303	7.541	7.900	1	60	139	5.281	4.993	1	41
Edit Difficulty (No Ex Vivo)	= 1 if disease affects cells that cannot be edited *ex vivo*; = 0 otherwise	442	0.314	0.465	0	1	303	0.000	0.000	0	0	139	1.000	0.000	1	1
Edit Difficulty (No Cell Rep)	= 1 if disease affects cells that do not self replicate; = 0 otherwise	442	0.172	0.378	0	1	303	0.000	0.000	0	0	139	0.547	0.500	0	1
Quarter	Quarter of focal CRISPR-Disease paper	442	2015q4	3.187	2013q1	2016q4	303	2015q4	3.277	2013q1	2016q4	139	2016q1	2.809	2013q1	2016q4
Quarters from First Pub	Difference in quarters from the focal to the first CRISPR-Disease paper in a Disease	442	2.394	3.345	0	15	303	2.779	3.439	0	15	139	1.554	2.971	0	14
Max Quarters from First Pub	Difference in quarters from the most recent to the first CRISPR-Disease paper in a Disease	442	4.541	4.272	0	15	303	5.304	4.231	0	15	139	2.878	3.885	0	14
Total Disease Pubs	Number of Disease papers published	442	506.380	662.740	2	2967	303	573.574	705.242	6	2967	139	359.907	532.585	2	2346
Total CRISPR-Disease Papers	Number of CRISPR-Disease papers published	442	1.382	0.981	1	8	303	1.426	1.023	1	7	139	1.288	0.878	1	8

Notes. Summary statistics by Disease-Quarter for CRISPR publications and authors in 228 Disease categories. CRISPR-Disease papers are the underlying population of the dataset and are defined as articles containing MeSH Terms for both the disease domain and CRISPR. Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.

Table 2. Counts of Authors and Papers by Disease Cell Edit Difficulty (*Ex Vivo* Editing) and Year

Total Number of CRISPR-Disease Authors by Edit Difficulty (No *Ex Vivo*)

	2013	2014	2015	2016	Total
Easy Diseases	27	232	730	1296	2285
Difficult Diseases	10	22	141	561	734
Total	37	254	871	1857	3019

Total Number and % of External CRISPR Specialists by Edit Difficulty (No *Ex Vivo*)

	2013	2014	2015	2016	Total
Easy Diseases	2	35	124	336	497
Difficult Diseases	0	6	47	186	239
Total	2	41	171	522	736

	2013	2014	2015	2016	Total
Easy Diseases	7.4%	15.1%	17.0%	25.9%	21.8%
Difficult Diseases	0.0%	27.3%	33.3%	33.2%	32.6%

Total Number of CRISPR-Disease Papers by Edit Difficulty (No *Ex Vivo*)

	2013	2014	2015	2016	Total
Easy Diseases	7	49	136	240	432
Difficult Diseases	4	7	30	138	179
Total	11	56	166	378	611

Notes. A CRISPR-Disease paper is defined as an article containing MeSH Terms for both the disease domain and CRISPR. CRISPR-Disease authors are the authors on these publications. External CRISPR Specialists are authors of CRISPR-Disease papers that published in CRISPR first. Year is when a CRISPR-Disease paper was published. Easy Diseases are those where the affected cell can be effectively edited *ex vivo*. Difficult Diseases are those where *ex vivo* editing is not available. Source: PubMed publications and MeSH Terms, 2013 – 2016.

Table 3. Disease Cell Edit Difficulty and Share of External CRISPR Specialists

	(1) OLS	(2) OLS	(3) OLS	(4) GLM, Logit	(5) GLM, Logit	(6) GLM, Logit
DV =	Share Ext. CRISPR Spec.	Share Ext. CRISPR Spec.	Share Ext. CRISPR Spec.	Share Ext. CRISPR Spec.	Share Ext. CRISPR Spec.	Share Ext. CRISPR Spec.
Edit Difficulty	0.0754*		0.1538***	0.3891*		0.7943***
(No *Ex Vivo*)	(0.0400)		(0.0483)	(0.2015)		(0.2198)
Edit Difficulty		-0.0291	-0.1459**		-0.1662	-0.7483**
(No Cell Rep)		(0.0506)	(0.0640)		(0.2532)	(0.3010)
Tot Disease Pubs	-0.0001***	-0.0001***	-0.0001***	-0.0007***	-0.0007***	-0.0007***
	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0002)	(0.0002)
Constant	0.0613	0.0931	0.0416	-15.2530***	-15.0626***	-15.4207***
	(0.0715)	(0.0624)	(0.0910)	(0.6577)	(0.7006)	(0.6997)
FE	Quarter, Age	Quarter, Age	Quarter, Age	Quarter, Age	Quarter, Age	Quarter, Age
Observations	442	442	442	442	442	442
Diseases	228	228	228	228	228	228
R^2	0.0853	0.0775	0.0991			
Log Likelihood	-132.9693	-134.8536	-129.6273	-204.7815	-205.7884	-202.8224

Standard errors in parentheses

Based on authors of all CRISPR-Disease papers, errors clustered by disease

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes. This table shows the difference in the share of external CRISPR specialists on CRISPR-disease papers by the difficulty of disease target cell editing using both OLS and GLM with Logit Link models. In diseases with cells that cannot be edited *ex vivo*, the share of external CRISPR specialist authors increases controlling for quarter, age, and disease attractiveness. The effect is stronger when also controlling for the target cell's ability to self-replicate. The difficulty of cell editing by disease can be measured by biological factors of the cells each disease primarily targets. Two key factors are (1) whether the cell can be edited *ex vivo* and (2) whether the cell can self-replicate. If the cell a disease targets cannot be edited *ex vivo* or if the cell does not self-replicate, then it will be harder and costlier (in resources) to edit. A CRISPR-disease paper is defined as an article containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR-disease paper has an external CRISPR background if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.

Table 4a. Share of External CRISPR Specialists in Subsequent Innovations by Disease Cell Edit Difficulty (*Ex Vivo* Editing) (Complete Dataset)

DV =	(1) OLS, FE Share Ext. CRISPR Spec.	(2) OLS, FE Share Ext. CRISPR Spec.	(3) GLM, Logit Share Ext. CRISPR Spec.	(4) GLM, Logit Share Ext. CRISPR Spec.
Edit Difficulty* Qtr. from First Pub	0.0442***		0.3331***	
	(0.0151)		(0.1292)	
Edit Difficulty* First Year (no first paper)		0.0294		0.0898
		(0.1152)		(0.8116)
Edit Difficulty* Second Year		0.1508		1.0440
		(0.1708)		(0.9814)
Edit Difficulty* Remaining Years		0.4737***		4.1928***
		(0.1562)		(1.3324)
Tot Disease Pubs	-0.0000	0.0000	-0.0009	-0.0003
	(0.0001)	(0.0001)	(0.0008)	(0.0009)
Constant	0.2345	0.1490	-18.4380***	-17.9954***
	(0.1780)	(0.1546)	(1.0464)	(1.0471)
FE	Quarter, Age, Disease	Quarter, Age, Disease	Age, Disease	Age, Disease
Observations	442	442	442	442
Diseases	228	228	228	228
R^2	0.1015	0.0966		
Log Likelihood	104.1781	102.9659	-115.3090	-115.2413

Standard errors in parentheses; Based on authors of all CRISPR-Disease papers, errors clustered by disease
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes. This table shows the difference in the share of external CRISPR specialists for subsequent CRISPR-disease papers by disease difficulty using both OLS and GLM with Logit Link models. As later subsequent CRISPR papers are published in difficult diseases as compared to easy diseases, the share of external CRISPR specialists authors increases controlling for disease, quarter, age, and disease attractiveness. The difficulty of the disease is measured by whether the targeted cell can be edited *ex vivo*. If the cell a disease targets cannot be edited *ex vivo*, then it will be harder and costlier (in resources) to edit. CRISPR-Disease papers are defined as articles containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR paper in a disease is considered an external CRISPR specialist if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.

Table 4b. Share of External CRISPR Specialists in Subsequent Innovations by Disease Cell Edit Difficulty (*Ex Vivo* Editing) (Diseases w/Many Papers)

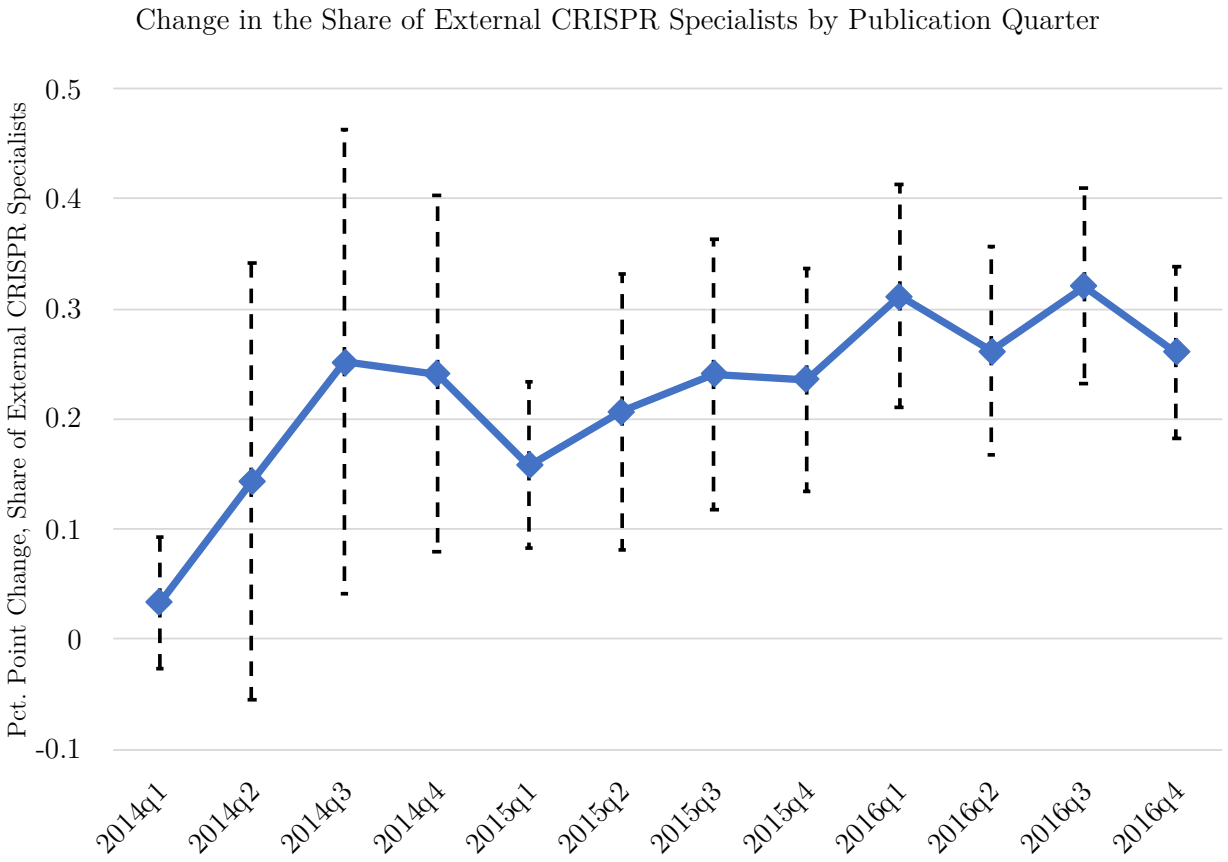
DV =	(1) OLS, FE Share Ext. CRISPR Spec.	(2) OLS, FE Share Ext. CRISPR Spec.	(3) GLM, Logit Share Ext. CRISPR Spec.	(4) GLM, Logit Share Ext. CRISPR Spec.
Edit Difficulty* Qtr from First Pub	0.0442***		0.3858**	
	(0.0153)		(0.1763)	
Edit Difficulty* First Year (no first paper)		0.0294		0.3494
		(0.1169)		(0.8458)
Edit Difficulty* Second Year		0.1508		1.3501
		(0.1732)		(1.1542)
Edit Difficulty* Remaining Years		0.4737***		4.0363***
		(0.1585)		(1.4842)
Tot Disease Pubs	-0.0000	0.0000	-0.0001	-0.0000
	(0.0001)	(0.0001)	(0.0012)	(0.0012)
Constant	0.1680	0.0895	-16.3330***	-17.4666***
	(0.1893)	(0.1658)	(2.8533)	(2.9530)
FE	Quarter, Age, Disease	Quarter, Age, Disease	Quarter, Age, Disease	Quarter, Age, Disease
Observations	307	307	307	307
Diseases	93	93	93	93
R^2	0.1015	0.0966		
Log Likelihood	16.4140	15.5721	-94.0049	-94.4182

Standard errors in parentheses; Based on authors of all CRISPR-Disease papers for diseases with more than one paper, errors clustered by disease; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes. This table shows the difference in the share of external CRISPR specialists for subsequent CRISPR-disease papers by disease difficulty using both OLS and GLM with Logit Link models but run only for diseases with multiple papers. As later subsequent CRISPR papers are published in difficult diseases as compared to easy diseases, the share of external CRISPR specialists authors increases controlling for disease, quarter, age, and disease attractiveness. The difficulty of the disease is measured by whether the targeted cell can be edited *ex vivo*. CRISPR-Disease papers are defined as articles containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR paper in a disease is considered an external CRISPR specialist if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.

10 Appendix

Figure A1. Quarter of Publication and Share of External CRISPR Specialists



Notes. This figure shows the change in the share of external CRISPR specialists by quarter of publication. Each point is the estimated coefficient and standard errors from an OLS model that regresses each quarter of publication from Q1 2014 – Q4 2016 on the share of external CRISPR specialists publishing CRISPR-disease papers. A CRISPR-disease paper is defined as an article containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR-disease paper has an external CRISPR background if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2014 – Q4 2016.

Table A1. Mean Differences in the Share of External CRISPR Specialists by Disease Cell Edit Difficulty (*Ex Vivo* Editing) for First and Subsequent Innovations

Difference in Means by Time Period and the Availability of *Ex Vivo* Editing

Share of External CRISPR Specialists	N (*Ex Vivo*)	N (No *Ex Vivo*)	*Ex Vivo*	No *Ex Vivo*	Diff	P-val
First Paper Only	138	90	0.23	0.33	0.10	0.05
In First Year - No First Paper	50	23	0.24	0.32	0.08	0.36
In Second Year - No First Paper	36	5	0.22	0.39	0.17	0.43
Remaining Years - No First Paper	13	6	0.21	0.47	0.26	0.24

Notes. This table shows the difference in means between the share of external CRISPR specialists for CRISPR papers in easy diseases versus difficult diseases based on whether the paper was the first or subsequent CRISPR paper in the disease and is the raw data for Figure 6. The share of external CRISPR specialists is higher on average for teams publishing CRISPR papers in difficult diseases for both the first and subsequent papers. The difficulty of the disease is measured by whether the targeted cell can be edited *ex vivo*. If the cell a disease targets cannot be edited *ex vivo*, then it will be harder and costlier (in resources) to edit. CRISPR-Disease papers are defined as articles containing MeSH Terms for both the disease domain and CRISPR. An author of a CRISPR paper in a disease is considered an external CRISPR specialist if he or she published in CRISPR first (and not in the disease). Source: PubMed publications and MeSH Terms, Q1 2013 – Q4 2016.