

# Governance of Biomedical Research Commons to Advance Clinical Translation: Lessons from the Mouse Model Community

*Tania Bubela, Rhiannon Adams, Shubha Chandrasekharan, Amrita Mishra, and Songyan Liu*

## INTRODUCTION

The translation of laboratory and clinical research into interventions that improve individual and population health is an iterative process with systemic directionality from basic research through preclinical research, clinical research, clinical implementation, and population health outcomes research (National Center for Advancing Translational Sciences 2015). Engaged in translation are patients and patient advocacy organizations, researchers from public and private sectors and from multiple disciplinary backgrounds, clinical practitioners, as well as a myriad of ancillary support professionals, from business executives, accountants, and marketing/sales staff to venture capitalists and public and philanthropic fund administrators to health and safety regulators.

New models of collaboration within this complex ecosystem are required to overcome waste and inefficiencies in current research and development (R&D)

Tania Bubela is Professor and Dean at the Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada. At the time this work was completed, she was Professor at the School of Public Health, University of Alberta, Edmonton, Alberta, Canada. Rhiannon Adams was a researcher with the School of Public Health, University of Alberta, Edmonton, Alberta, Canada. Shubha Chandrasekharan is Assistant Research Professor, Duke Global Health Institute, Duke University, Durham, North Carolina, USA. Amrita Mishra was a researcher at the School of Public Health, University of Alberta, Edmonton, Alberta, Canada. Songyan Liu is a researcher in bioinformatics at the Research Institute in Oncology and Hematology, University of Manitoba and CancerCare Manitoba, Winnipeg, Manitoba, Canada.

Many individuals have participated and supported this research over the years. Most prominently, with respect to data and patent analyses, have been Dr. Andreas Strotmann, Dr. Sean Morrison, and Mr. Mark Bieber. Expert advice has been provided most notably by Dr. Lauryl Nutter, Dr. Ann Flenmiken, Mr. David Einhorn, Dr. Edna Einsiedel, and Dr. Paul Schofield as well as the nearly 100 interviewees, internationally, who generously gave their time to inform and comment on our research. Research assistants and staff have additionally included Dr. Cami Ryan, Ms. Noelle Orton, Mr. Amir Reshef, and Ms. Lesley Dacks. Funding for the research has been provided by Genome Canada and Genome Prairie for the NorCOMM I Project (PI: Dr. Geoff Hicks; co-PI Dr. Janet Rossant); Genome Canada and the Ontario Genomics Institute for the NorCOMM II (Project Co-Lead Investigators: Colin McKerlie and Steve Brown); the Canadian Stem Cell Network funding to Dr. Tania Bubela.

within pharmaceutical or larger biotechnology companies (Hunter 2014; Munos 2009). Despite substantial increases in R&D investments to US\$50 billion per year, the number of new drugs approved annually in the United States has remained constant over the past 60 years – the US Food and Drug Administration (FDA) has approved approximately 1200 new drugs (Munos 2009). The high cost of therapies is in part driven by the cost of failures in clinical development; only 15.3 percent of drugs traverse the pipeline from Phase 1 to market authorization for lead indications, a percentage that drops to 10.4 percent for all indications (Hay et al. 2014). Despite billions of dollars of investment, new cancer drugs developed between 2002 and 2014 have produced median gains in progression-free survival of only 2.1 months and median gains in overall survival of only 2.5 months (Fojo et al. 2014). This expenditure of time, money, and resources on marginal therapeutic benefits is “promoting a me-too mentality that is stifling innovation and creativity” (Fojo et al. 2014).

Some argue that precompetitive research partnerships can overcome the innovation and creativity gap by drawing on the respective strengths of different sectors to facilitate R&D of new therapies and diagnostics in the translational environment. Research commons can provide infrastructure to support such precompetitive collaborations between academia, government, industry, nongovernmental organizations, and patients (Bubela et al. 2012a; Edwards et al. 2009; Friend 2010). Precompetitive collaborations can facilitate sharing of data and materials without limiting the ability of commercial actors to appropriate knowledge that is closer to practical application. They aim to raise the knowledge levels for all R&D actors to avoid duplicative research while facilitating replication for validation, and to promote the use of standard research tools and methods.

Resources governed as research commons may be informational (e.g., databases), material (e.g., biorepositories and biobanks), or a combination of the two. Commons governance of information is especially important if one accepts the argument that the principal output of clinical translation is information, defined as including “information about the coordinated set of materials, practices, and constraints needed to safely unlock the therapeutic or preventive activities of drugs, biologics, and diagnostics” (Kimmelman and London 2015). Commons governance of materials allows researchers to share quality-controlled, standardized research tools (Schofield et al. 2009). Most importantly, the value of commons governance increases as more people use the shared resources. Pooled resources must therefore “be managed to facilitate use, but also re-contribution from the user community, creating a feedback loop between withdrawal, value-added research, and deposit” (Bubela et al. 2012b: 107).

Here we focus on one of the most established biomedical research commons – the mouse commons – that is composed of both biorepositories and databases, which we collectively refer to as archives (Einhorn and Heimes 2009; Schofield et al. 2009). The mouse model community established this commons in response to the challenge of coordinating the sharing of data and research reagents generated by high-

throughput ‘omics’ technologies (Bubela et al. 2012a). We ask what lessons may be learned from this model organism community for others seeking to develop governance structures that will sustain long-term access to international research resources. We address the history and development of the mouse commons and key challenges in its development using Elinor Ostrom’s Institutional Analysis and Development (IAD) Framework as modified for the study of knowledge commons (Frischmann, Madison, and Strandburg 2014). We focus on the role of formal intellectual property rights and contracts in promoting or impeding research commons as well as on rules development and implementation relevant to incentivizing, dis-incentivizing, and sanctioning participation in the commons. Finally, we discuss governance mechanisms that promote the long-term sustainability of the commons, especially with respect to the high costs associated with sustaining archives that house the resources.

Our analysis is based on a decade of work with the mouse model research community – we conducted nearly 100 interviews with a variety of international stakeholders, including managers of mouse commons infrastructure, high-throughput resource generators, funders, and research users (Bubela et al. 2012a; Bubela, Guebert, and Mishra 2015a; Mishra and Bubela 2014; Mishra, Schofield, and Bubela 2016). In addition, we participated in community-level infrastructure planning workshops, analyzed the patent landscape up to 2007, and analyzed contractual agreements that govern sharing by repositories and databases up to 2013.

#### 10.1 APPLICATION OF THE IAD FRAMEWORK TO THE MOUSE MODEL RESEARCH COMMONS

Elinor Ostrom’s IAD Framework enables the systematic study of the governance of commons, whether these be natural resource commons such as forests of fisheries (Ostrom 1990, 2005) or research commons where the goal is to share research reagents, know-how, and data (Dedeurwaerdere 2010a, 2010b; Hess and Ostrom 2006, 2007). Since research commons are a type of knowledge commons (knowledge in this context refers to “a broad set of intellectual and cultural resources” (Frischmann et al. 2014), the adapted knowledge commons framework provides the appropriate mechanism to study commons governance that aims to achieve broad-based participation and sharing of data and materials for research (Frischmann et al. 2014). This sharing is dependent on the accepted norms and behaviors of the research community – in this case, researchers who use mouse models to study human disease. These norms and behaviors may be impacted by the heterogeneity of the community, which creates challenges in managing differing sets of norms and behaviors about sharing versus withholding of data and materials (Bubela et al. 2012a). The main concern of individual researchers that contributes to the low 35 percent sharing rate for mice is the “free-rider” problem – that is, the prevention of users benefiting from the commons without contributing to it (Schofield et al. 2009). A further problem is the simple logistical burden of sharing

mouse-related research materials with other researchers, one overcome by the one-stop deposit to a repository that then takes on the burden of onward distribution to the research community (Einhorn and Heimes 2009; Schofield et al. 2009). In addition, current trends toward commercialization of publicly funded outputs from research institutions have led to a plethora of intellectual property rights (IPRs), mostly patents, over mouse-related research materials. Ironically, the development of the mouse research commons is, in part, a reaction to commercialization policies of the same governments, funders, and research institutions that now promote the research commons (Caulfield et al. 2012; Popp Berman 2012).

A successful commons requires rules and governance structures to overcome some of these challenges (Ostrom 2005). Rules, ideally developed with the participation of community members; need to incentivize creation of the commons; contribution to the commons; use of the commons; and, most importantly, re-contribution of value-added data and materials to the commons. Rules need to provide a graduated system of sanctions for noncompliance; but in a complex ecosystem, debate occurs as to which entities should develop and enforce these sanctions and what mechanisms to employ for conflict resolution (Ostrom and Hess 2007). In a research commons, the hierarchy of rules usually ranges from national-level laws that govern IPRs, regulatory approval processes, and animal welfare to policies and guidelines for contractual arrangements, funding, and collaborative research. At their simplest, commons governance rules may include norms and practices within a community around citation, attribution, reciprocity and sharing, and form and timing of publication. In addition, governance models need to ensure the long-term sustainability of the commons, while remaining versatile and responsive to technological change (Mishra et al. 2016).

In the following sections, we first describe the background environment and history of the mouse model research commons, including its heterogeneity. We then discuss IPRs as impediments to the development of the mouse commons, the rules that have been put in place to incentivize participation in the commons, and the governance structures to support the long-term sustainability of the commons.

## 10.2 THE COMMONS ENVIRONMENT FOR MOUSE MODEL RESEARCH

The mouse is the quintessential biomedical research tool. With its genetic similarities to humans, it is an efficient and effective model system to assist researchers in understanding links between genes and disease states (Kim, Shin, and Seong 2010). Knockout mice are genetically manipulated to lack one or both copies of a specific gene (or portion thereof), with corresponding impact on the protein the gene encodes. Knockin mice carry engineered copies of genes that may still be functional. For example, knockin mice may be genetically manipulated to carry human genes that make them susceptible to cancer or that “humanize” parts of their immune system for drug testing. The scientific significance of mouse models was recognized

by the 2007 Nobel Prize in Physiology or Medicine awarded to Dr. Mario R. Capecchi and Sir Martin J. Evans and Dr. Oliver Smithies for their early work on knockout mice.

In the 1980s, the generation of genetically modified mice was technically challenging, required highly skilled personnel, and it was costly in terms of time and research funds. The mice themselves, as mutants, were physiologically fragile and difficult to breed. Generation of mutant mouse models that were the bedrock of biomedical research was therefore beyond the skills and resources of many research laboratories, and those laboratories with model generation capabilities were reluctant to share these valuable resources. Since public funding agencies invested considerable resources into the generation of the mouse models, it became imperative to develop infrastructure capable of receiving strains from generating laboratories and distributing the strains to other users in the research community.

The Jackson Laboratory (JAX) was the first international mouse repository. It was created in the 1930s to distribute community-generated mice in addition to its own strains (Einhorn and Heimes 2009). JAX became a frozen embryo repository in 1979 and is now the world's largest repository of live research mouse strains (Jackson Laboratory 2014). JAX and other repositories promote and facilitate access to research tools and are at the epicenter of the mouse research commons. Such repositories accept deposit of mouse-related resources from individual research laboratories and then archive and distribute the resources using conditions of use or material transfer agreements (MTAs) that promote research use (Mishra and Bubela 2014). More recently, the commons resources have come to include mouse embryonic stem cell lines (mESCs) and gametes (mainly sperm), which are more efficient to store and distribute and can be used to generate live mice. They also now collect derivative cell lines, vectors for genetic manipulation, and associated genotyping and phenotyping data (Brown and Moore 2012a, 2012b; Collins et al. 2007; Skarnes et al. 2011). Collectively, these materials and associated data comprise a comprehensive mutant mouse resource.

Despite this robust international sharing infrastructure, only approximately 35 percent of generated mouse strains are made available to the research community (Schofield et al. 2009). Partly in response to this statistic, the international community launched the International Knockout Mouse Consortium (IKMC) in 2007 (Collins et al. 2007; Skarnes et al. 2011). The IKMC is generating a mutant resource for all protein-coding mouse genes that can be archived and distributed by repositories affiliated with the IKMC (Bradley et al. 2012). It has the added benefit, as a high-throughput pipeline, of enhancing efficiency and reducing the costs of developing mouse models, which formerly were funded through individual research grants. Those individually developed mouse models also may not have been of high quality, were non-standardized with respect to background strain of the mouse, and may or may not have been shared.

A second international consortium, the International Mouse Phenotyping Consortium (IMPC) was established in 2011 to add standardized phenotyping data to the IKMC resource, thereby generating an encyclopedia of mammalian gene function (Brown and Moore 2012a, 2012b). The user community could nominate genes to be prioritized for both production (IKMC) and phenotyping (IMPC). The IMPC is expanding to include additional data from secondary phenotypers, who will contribute additional specialized screens, and from experienced end users, who will re-contribute tertiary phenotyping data, working in collaboration with the IMPC centers (Adams et al. 2013). This ability to accept data from secondary phenotypers creates the network effect (the re-contribution of value-added resources) that promotes research commons.

The IKMC and the IMPC are both international consortia of mouse genetics centers, supported by national and regional funding bodies in North America, Europe, and Asia-Pacific. Each relies on government-funded infrastructure for archiving and sharing mouse strains and associated data. For example, IKMC resources are available from the Knockout Mouse Project (KOMP) Repository housed at University of California, Davis ([www.komp.org/](http://www.komp.org/)) and the European Mouse Mutant Cell Repository (EuMMCR) at Helmholtz Zentrum in Munich, Germany ([www.eummc.org/](http://www.eummc.org/)). Phenotyping data generated by the primary IMPC centers are processed, housed, and made available via the NIH-funded KOMP Data Coordination Center and the European Commission–funded International Data Coordination Center. These data coordination centers provide common semantics for comparing and integrating imaging, text-based, and numerical data (Mallon et al. 2012; Ringwald and Eppig 2011).

In summary, the mouse research commons is composed of (1) individual researchers who produce and/or use mouse models; (2) repositories and databases where individual researchers may deposit or access mouse models or data, such as JAX and Mouse Genome Informatics (MGI), the international database resource for the laboratory mouse ([www.informatics.jax.org](http://www.informatics.jax.org/)); (3) the high-throughput production centers for community resources that form the IKMC and IMPC consortia; (4) the high-level end users that contribute data to the IMPC; and (5) the national and regional funders that support the commons. The functioning of this commons is premised on rules and governance structures that promote the sharing of mouse research tools and associated data for biomedical research.

### 10.3 LEGAL ISSUES THAT IMPACT THE COMMONS ENVIRONMENT

#### 10.3.1 *Historical Context*

In creating the mouse model research commons, the community confronted a number of well-known controversies over access to mice as research tools. Researchers at Harvard University created the OncoMouse, a knockin mouse with

a predisposition to develop cancer. The OncoMouse strains and associated methods were covered by broad patents that were exclusively licensed to DuPont, which funded the research (Murray 2010). DuPont imposed restrictive licensing terms that were contrary to community norms for sharing valuable mouse strains that were emerging in the 1980s. The licenses restricted third-party distribution of novel strains developed in individual laboratories from oncomice. They required annual disclosure of research findings using the mice and imposed reach-through rights on future discoveries that entitled DuPont to “a percentage share in any sales of proceeds from a product or process developed using an OncoMouse, even though the mice would not be incorporated into the end product” (Murray 2010: 362).

Community resistance to these restrictions necessitated a negotiated compromise between DuPont and the National Institutes for Health (NIH), which signed a memorandum of understanding (MOU) in 1999. The MOU enabled academic researchers to share strains under simplified conditions of use. These conditions of use no longer included reporting requirements or reach-through-rights. JAX and other public repositories then made the strains widely accessible to academic institutions that had funding agreements with the Public Health Service of the US Department of Health and Human Services (DHHS). Researchers at institutions not so funded, including those outside of the United States, were advised to seek a license for use from DuPont (Mishra and Bubela 2014).

The NIH also intervened to enable access to *cre-lox* technology that was developed in DuPont’s life sciences division in 1987. *Cre-lox* technology generates conditional mutants with genetic modifications expressed only in specific tissues (Sauer and Henderson 1988). Restrictive licensing agreements also limited access to this patented, powerful research tool for studying gene function. In 1998, NIH negotiated an MOU to allow JAX and institutions with funding agreements with the Public Health Service of the DHHS to distribute and share *cre-lox* mice, subject only to simple conditions on use. In light of these two NIH negotiated MOUs, follow-on research, measured through citations, increased (Murray et al. 2009). The MOUs also encouraged new authors at a greater diversity of institutions to conduct research using the mouse technology in a broader range of fields (Murray et al. 2009). Thus the development of institutional mechanisms for generation and distribution of mouse strains promoted sharing of these valuable research tools, one of the goals of a research commons.

These cases, however, also suggest that the seeking and rigorous enforcement of intellectual property rights may impede the creation and successful functioning of research commons. Since the 1980s, government and funding agency policies have incentivized the seeking of formal intellectual property rights by researchers receiving public funds, while promoting the sharing of data and research materials (Caulfield et al. 2012). The mouse-model research community was no exception, and our analysis of patents covering mouse-related research reagents, described in Section 10.3.2, demonstrates the extent to which incentives for academics to patent research tools have been effective. Such patents, with a few exceptions, are, in a

practical sense, largely worthless in that they generate little to no revenue for the patent holder relative to the cost of patent acquisition and maintenance. Commons are a more effective mediator of exchanges of research tools because they reduce transaction costs associated with sharing, especially for research tools that are largely generated using public research funds and then used by other publicly funded researchers. The premium that may be charged for patented research tools makes limited sense from a social perspective because it represents a research tax from others within the same community, with public research dollars simply flowing from one institution to another. Indeed, our interviewees clearly stated the uselessness of patenting mouse-related research tools and expressed dissatisfaction with the entities that mediated such practices, namely, their technology transfer offices (Bubela et al. 2012a).

### 10.3.2 *Do Intellectual Property Rights Impede the Creation and Functioning of Research Commons?*

In analyzing the mouse patent landscape, we asked whether a legacy of basic research patents created under current laws and practices hindered the establishment of public sector commons infrastructures. Specifically, we (1) explored the characteristics of mouse genes that had been patented prior to September 2007 compared to a control set of unpatented mouse genes and (2) compared research activity based on patented mouse genes with research activity based on non-patented mouse genes. Briefly, on September 27, 2007, we searched the Thomson database, Delphion, for granted US patents that (1) had variants of the terms “mouse” or “mammal” in claims, (2) matched a modified Ade/Cook-Deegan Algorithm for RNA/DNA patents, and (3) did not have variants of “plant” in the claims. The search identified 7179 granted US patents, from which we extracted standard patent information from the database and which we then read and coded to identify the 1144 patents that claimed gene sequences, by SEQ ID (DNA/RNA/amino acid) or a gene name prior to 1996. In the Appendix, we describe our methods for analyzing our patent data set.

Contrary to the beliefs of many of our interviewees in the mouse model community, there was considerable patenting of mouse-related research reagents, especially during the heyday of “gene” patenting in the late 1990s through 2001 (Bubela et al. 2012a; Carbone et al. 2010; Cook-Deegan and Heaney 2010a, 2010b). Not only were such patents sought and granted, but most were maintained over the course of their patent terms (Figure 10.1).

The majority of mouse DNA and mouse patents were held by public and private universities and other nonprofit entities, which reflects a considerable investment by those organizations in maintaining a patent portfolio covering low-value research-related subject matter (Figure 10.2). By contrast, pharmaceutical and biotechnology companies were more strategic in their patent filing and maintenance. These companies dominated ownership of broad DNA claims that overlapped mouse DNA but were



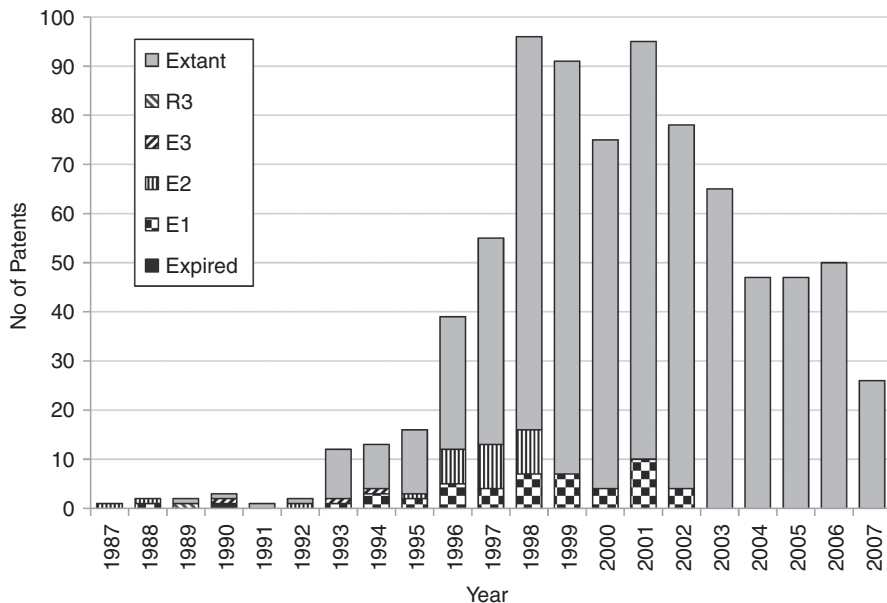


FIGURE 10.1 Maintenance status and filing date of 816 mouse gene patents (US) between 1987 and 2007. Maintenance codes are assigned by the USPTO: E1 – expired 4 years after the initial issue date; E2 – expired 8 years after the initial issue date; E3 – expired 12 years after the initial issue date; R3 reinstated after E3.

worded to include DNA from other mammals. They also dominated ownership of patents that claimed cell lines. Mice as research tools were claimed by all sectors.

To answer our question on the impact of patents on follow-on mouse model research, we identified each of 951 patented genes based on gene name or a blast analysis of the DNA sequences listed in the patents (see Appendix for a detailed explanation of the analysis). Using Online Mendelian Inheritance in Man (OMIM), the Mouse Genome Database (MGD), and the Mouse Genome and Informatics (MGI) databases, we detailed the characteristics of patented mouse genes. We compared these to a set of 1354 randomly selected unpatented genes identified from the same databases. The 951 patented genes were twice as likely to be listed in the OMIM database, 30 percent more likely to have a human orthologue, and nearly three times as likely to have a defined phenotype (Table 10.1). These are all indicators of research importance to human health – in other words, patented genes had more indicators relevant to research on the genetics of human diseases than our sample of unpatented genes.

We then examined the impact of gene patents on research outputs. Of the 108 mouse genes with greater than 100 associated research publications in PubMed, 86 (79.6%) were patented (Figure 10.3). Since number of publications is a metric for

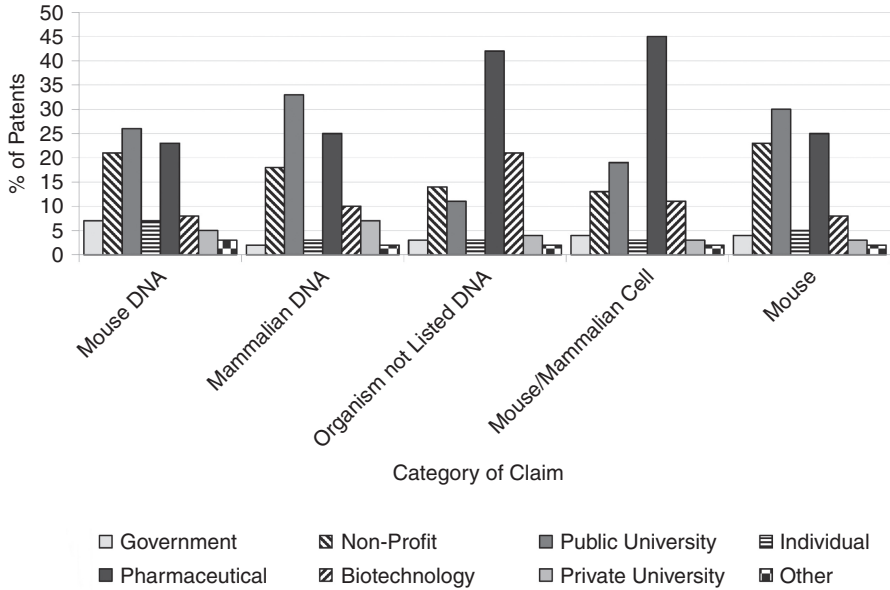


FIGURE 10.2 Percentage of 7179 US granted patents with claims that fell into five categories of mouse-related subject matter by type of assignee. Note that claims within one patent could fall within multiple categories of subject matter.

TABLE 10.1 Characteristics of patented versus unpatented genes

	Patented Genes (n=951)	Unpatented Genes (n=1397)
Gene Traps	600	1354
Genes on Targeting List	86.3% (821)	76.4% (1068)
Total Number of Targeting Designs	632	808
OMIM ID	64.7% (615)	39.4% (550)
OMIM Description	20.0% (190)	10.0% (141)
Gene Phenotypes	51.0% (485)	19.0% (266)
Human Orthologue	99.4% (945)	61.3% (856)

research intensity, the high level of research by groups of authors in addition to the named inventors likely represents broad community ignorance or knowing infringement of the identified patent portfolio (see also Walsh et al. 2005a, 2005b, 2007). Our interviews confirmed that the research community pays little attention to patents and/or believes that there is a broad-based research exemption. In the United States, especially after the law was clarified in *Madey v Duke*

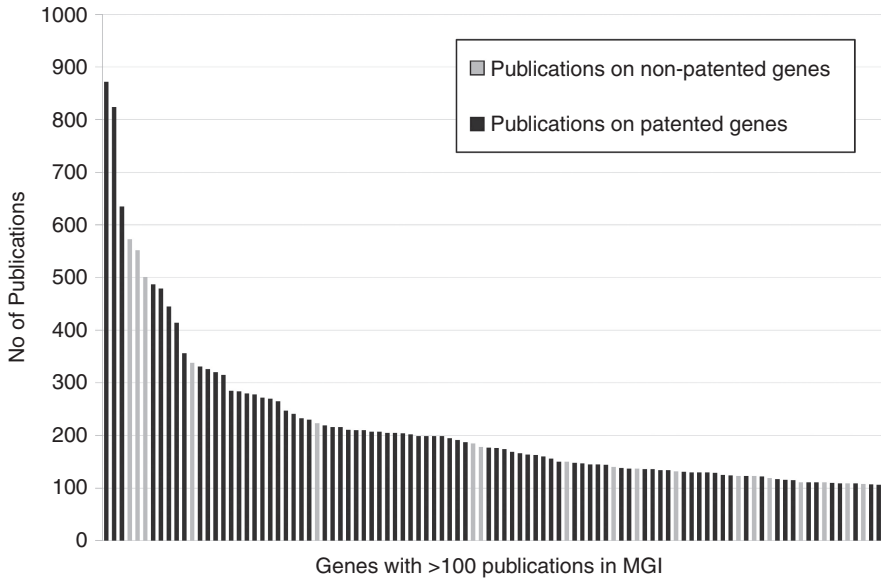


FIGURE 10.3 Genes with more than 100 associated publications in the Mouse Informatics and Genomics (MGI) database illustrating genes that are patented in the United States (black) versus those that are not patented (grey). Each bar represents one gene.

*University*,<sup>1</sup> there is a very limited research exemption in law. Nevertheless, a de facto research exemption operates because most preclinical researchers are simply not worth suing (Cook-Deegan and Heaney 2010a; Heaney et al. 2009). This further supports our contention that this patent portfolio over mouse-related research tools is of limited commercial value and merely muddies the waters for any researchers who attempt in good faith to ascertain their freedom to operate.

This conclusion could be encouraging even though it is based on broadly infringing research activity. However, our analysis further indicated that the *rate* of publications on mouse genes, based on annual rate of change, began to decline one year after a patent was granted. Publication rates for patented genes significantly decreased three years post-patent grant compared to the same time period prior to patent grant (Table 10.A2). The publication rates of our comparator set of unpatented genes remained constant over time, with no significant fluctuations in publication rate (Table 10.A4). Indeed, the publication rate for patented genes was significantly reduced compared to non-patented control genes three years post-patent grant. While the publication rate for patented genes increased to pre-patent publication levels five to seven years post-patent grant, it remained lower than publication rates on the comparator set of non-patented genes – a reversal from the pre-patent period (Figure 10.4; Tables 10.A6 and 10.A7). Prior to patent grant,

<sup>1</sup> 307 F.3d 1351 (Fed. Cir. 2002).

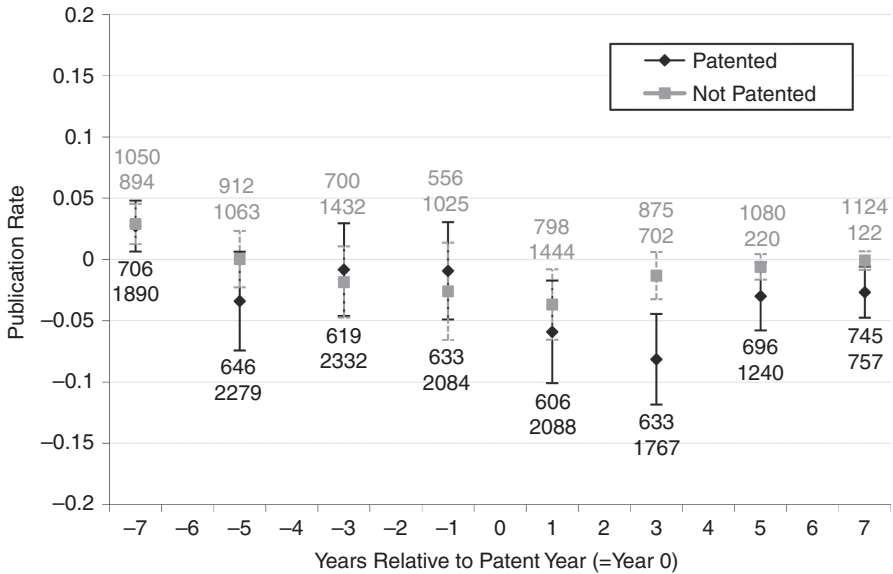


FIGURE 10.4 Publication rates of patents versus unpatented genes relative to the year of patent issue. See Appendix for methods of calculation.

publication rates for patented genes were non-significantly higher than for non-patented genes, also reflected in the fact that patented genes were among those with the highest numbers of associated publications.

For citations, which indicate follow-on research, the citation rate of patented genes declined significantly three years post-patent grant compared to similar time periods prior to patent grant (Table 10.A3); however, this trend likely reflects expected patterns of citation that rise to a peak and then decline over time – unpatented genes demonstrated the same trend of declining citation rates over time (Table 10.A5). Citation rates to publications over unpatented genes were significantly higher than publications on patented genes in six out of eight time periods (Table 10.A7). Taken together, our results imply that patenting had a negative impact on the contribution of knowledge in the scientific literature on mouse genes relevant to human health.

Our analyses imply that patenting of mouse genes had a discernible negative impact on follow-on research that used mouse models to study human disease. However, in our view, composition of matter patents were not the major impediment to the development of a mouse research commons, especially to that portion of the commons driven by high-throughput initiatives to build an international knock-out mouse infrastructure. In this context, the creation of a high-throughput pipeline to generate standardized research reagents and associated data requires the aggregation of platform technologies used to generate the resource. The greatest

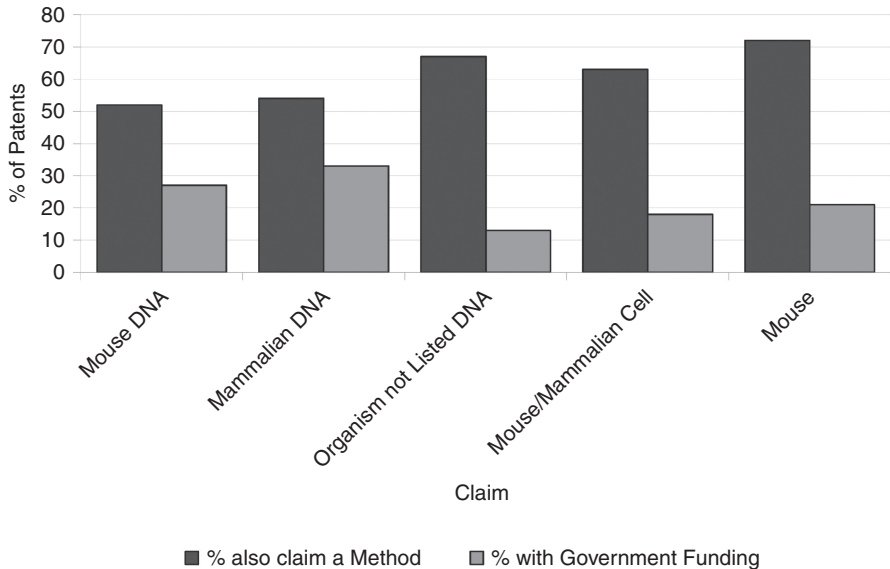


FIGURE 10.5 Percentage of 7179 US granted patents with claims that fell into five categories of mouse-related subject matter that indicated government funding and that claimed a method.

impediment to such aggregation was therefore the patenting of broad-based methods used to generate the resource. While we found that the majority of patents that claimed mouse-related compositions of matter also claimed a method for generating those materials (Figure 10.5), the most problematic 105 patents covered methods to generate the resource, held by a combination of industry and research institutions (Table 10.2).

Our conclusions on the impact of patenting of mouse-related research reagents may be summarized as follows. Our analysis identified a large number of overlapping patents on both compositions of matter and methods. Patents over mouse genes were associated with other indicators of research indicating relevance to genetics of human diseases. Most DNA patents were held by the public sector, while most cell lines and mouse-model patents were held by the private sector, reflecting their respective value to commercial and noncommercial sectors. We question the utility and expense to research institutions in maintaining this low-value patent portfolio, and this finding is likely indicative of increasing incentives to patent research outputs combined with a lack of resources within technology transfer offices to manage (i.e., prune) patent portfolios over their lifespan. Since preclinical mouse-related research is far from clinical application (timelines from Phase 1 clinical trials in humans to market authorization range from 10 to 14 years for small molecule drugs (Hay et al. 2014)), such patents are unlikely to generate

TABLE 10.2 *Mouse-related granted US methods patents by assignee as of 2007. The most problematic of the 105 methods patents covered BAC, positive/negative selection, FLP/FRT recombinase, isogenic DNA, recombinering, electroporation, PhiC31, Cryopreservation, Gateway technology, Cre-lox, inverse PCR, Vector construction, and homologous recombination. See table footnotes for company mergers and name changes current to 2016.*

Assignee	No. of Patents
None	9
Lexicon Genetics, Inc.	9
Institut Pasteur	7
Regeneron Pharmaceuticals, Inc.	6
University of Utah Research Foundation	6
The Salk Institute for Biological Studies	5
The Regents of the University of California	4
Invitrogen Corporation <sup>1</sup>	4
Genpharm International, Inc. <sup>2</sup>	4
Amgen, Inc.	3
Wisconsin Alumni Research Foundation	3
Artemis Pharmaceuticals, GMBH <sup>3</sup>	2
Deltagen, Inc.	2
Europaisches Laboratorium für Molekularbiologie (EMBL)	2
Idec Pharmaceuticals Corporation <sup>4</sup>	2
Institut Curie	2
Roche Diagnostics, GMBH	2
The Institute of Physical and Chemical Research	2
The Jackson Laboratory	2
The University of Edinburgh	2
Biogen Idec, Inc. <sup>5</sup>	2
Centre National de la Recherche Scientifique	2

<sup>1</sup> Acquired by Thermo Fisher Scientific, Inc.

<sup>2</sup> Acquired by Bristol-Myers Squibb Company.

<sup>3</sup> Renamed Taconic Biosciences, Inc.

<sup>4</sup> Merged with Biogen, Inc.

<sup>5</sup> Renamed Biogen, Inc.

revenues from clinically available therapies. In addition, the patent portfolio covers preclinical *research* tools that are widely infringed, making it unlikely that a university will generate any licensing revenue. Nevertheless, any revenue is most likely to come from sister research institutions (the main users of mouse models for research), comprising a tax on public research resources.

While it appears that most researchers ignore this patent landscape, there are measurable impacts on publication and citation rates. Based on our key-informant interviews, the biggest impact lies in cultural shifts toward proprietization within the research community, which negatively impacts efforts to incentivize sharing of research reagents and associated data. Finally, broad-based methods patents may impede the generation of high-throughput resources, designed to increase research efficiency and create a research commons. We discuss the impact, based on interviews, of methods patents on high-throughput resource generation and the impediments posed to rules development and governance in the next sections.

#### 10.4 RULES TO INCENTIVIZE AND FACILITATE PARTICIPATION IN THE COMMONS

Rules to incentivize and facilitate participation in the mouse research commons are promulgated by varied actors. As stated earlier, we use Ostrom's definition of rules-in-use as consisting of "shared normative understandings of what a participant in a position must, must not, or may do in a particular action situation, backed by at least a minimal sanctioning ability for noncompliance" (Ostrom and Hess 2007: 50). Ostrom outlines a hierarchy of rules as including formal laws; constitutional laws; policies and guidelines; and informal rules, community norms, and practices. In the previous section, we outlined the impacts of one set of formal laws – national intellectual property laws. Such generic laws are often out of sync with new capabilities, community norms, and technological advances (e.g., the development of technology platforms for high-throughput generation of research reagents). Other relevant formal laws include those associated with animal welfare standards and preclinical research requirements for regulatory approvals to advance to clinical trials. The former are closely tied to the development of the resource-generation projects, which aim to avoid duplication in the generation of live-animal models, thereby reducing the number of mice used in research (Russell and Burch 2014).

Earlier we also suggested that the policies and guidelines of funding agencies and research institutions that incentivize commercialization activities, including the seeking intellectual property rights over research outputs and partnerships with industry, may dis-incentivize participation in the commons. These policies result in an increase in secrecy and data withholding that may be contrary to the goals of open science (Walsh et al. 2005a, 2005b). Other policies and guidelines from funders, however, are broadly supportive of a research commons. Indeed, public funders internationally supported the resource initiatives of the IKMC and the IMPC, as well as existing biorepositories that are central to the mouse research commons, such as JAX.

Beyond funding, however, policies and guidelines supportive of the commons need to incentivize contributions to the commons, use of the commons, and activities that add value to the commons. These policies and guidelines codify

community norms and practices for the sharing of mouse-related research tools and data. Many such guidelines exist. For example, most funding agencies require the deposit of publications and associated data into public databases (Van Noorden 2012; Whitfield 2011), and the NIH even has guidelines on the sharing of biosources developed using NIH funds (Field et al. 2009).<sup>2</sup> There is, therefore, no shortage of policies and guidelines to promote sharing, but there is limited enforcement of these policies and guidelines, especially as they relate to deposit of mouse-related biomaterials (Schofield et al. 2009). Enforcement could include a denial of funding to researchers who cannot provide evidence of deposit for research tools into a recognized archive as is being contemplated by the NIH for researchers who cannot provide evidence of deposit of publications into open access archives within one year of publication (Grant 2012). Similarly, journals could deny publication without such evidence; some journals already require, for example, an accession number (evidence of deposit into GenBank) for sequence data prior to publication.<sup>3</sup> In terms of incentives, researchers are driven by publications and citations to those publications. The expansion of accession numbers to mouse-related research tools could provide a mechanism for attribution of effort and a means for citation to resources generated in addition to publications. These could then be recognized by research institutions in respect to assessments of researcher merit, tenure, and promotion. In other words, mechanisms for incentivization and enforcement exist but are yet to be fully implemented (Schofield et al. 2009).

In addition to policies and guidelines, MTAs and Data Transfer Agreements (DTAs) mediate the exchange of materials and data, respectively. In our case study, MTAs covered exchanges for mouse-related research reagents. As evidenced by the private sector MTAs initiated by DuPont in the 1980s, MTAs may be used to extend and protect proprietary interests over both patented and unpatented technologies. MTAs are almost universally despised by academic researchers in relation to research tools (Bubela et al. 2015a). Our interviews suggest that most academic researchers find MTAs problematic in their complexity. When mediating exchanges among collaborating researchers at different institutions, institutions insert overly onerous terms that delay negotiations and the transfer of materials (Mishra and Bubela 2014). This is especially problematic for low-value research tools in the precompetitive environment that are far from clinical application.

As types of licensing agreements (contracts that grant a permission to use), MTAs may also be used to embody the policies of funders and practices of the community with respect to the creation, maintenance, and functioning of a commons (Bubela et al. 2015a; Mishra and Bubela 2014). In other words, MTAs may be structured in such

<sup>2</sup> National Institutes of Health. 1999. Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources. 64 *Federal Register* 72090. Retrieved from [http://grants.nih.gov/grants/intell-property\\_64FR72090.pdf](http://grants.nih.gov/grants/intell-property_64FR72090.pdf)

<sup>3</sup> NCBI GenBank. 2015. How to Submit Data to GenBank. Retrieved from [www.ncbi.nlm.nih.gov/genbank/submit/](http://www.ncbi.nlm.nih.gov/genbank/submit/)



a way that they simplify and promote sharing of research reagents rather than imposing limits on use of the materials and generating revenue and other benefits. We analyzed the extent to which the MTAs that cover mouse research tools embody sharing policies in Mishra and Bubela (2014) and found that the MTAs used to distribute mouse-related resources are generally supportive of the creation of a mouse research commons, at least among nonprofit researchers and their institutions. Since MTAs are also part of the broader governance structure of a research commons, we considered the role of repositories as mediators of exchanges between resource generators (whether individual laboratories or high-throughput resource generation programs such as members of IKMC).

In 1995, the National Institutes of Health published the Universal Biological Materials Agreement (UBMTA) and a Simple Letter Agreement for the Transfer of Non-Proprietary Biological Material (SLA) as general models for transfer of biological materials.<sup>4</sup> For research tools, the NIH and the Association of University Technology Managers (AUTM) recommend use of the SLA because the UBMTA is more complex and provides additional protections for patented materials (though it may also be used for unpatented materials). The KOMP repository uses an MTA largely based on the UBMTA for the distribution of its mouse resources, except that, under the terms of its funding, it additionally enables distribution to commercial entities – this additional requirement to distribute to industry means the KOMP repository cannot use the SLA (Mishra and Bubela 2014).

MTAs are used by repositories, central actors within research commons, to govern both the terms under which resources are deposited into the repository and the terms under which those resources are distributed to users. Issues arise when the deposit terms limit or otherwise impact the distribution terms. Variable terms in deposit MTAs, such as differential restrictions on commercial versus noncommercial research, need to be tracked, attached to the materials as metadata, and transferred to the distribution MTA. The operations of the repositories, and accordingly the operation of the commons, would be made more fluid by consistent, simplified terms governing both deposit and distribution. Highly variant MTAs create an

<sup>4</sup> The SLA provides that the material to be transferred (1) remains the property of the provider; (2) will not be used in human subjects; (3) will be used only for teaching and nonprofit research purposes; (4) will not be further distributed to third parties without permission from the provider; (5) the material will be acknowledged in publications; (6) standard warranty and liability provisions; (7) use of materials will be in compliance with laws and regulations; and (8) the material is provided at no cost beyond those associated with preparation and distribution (Retrieved from [www.autm.net/resources-surveys/material-transfer-agreements/nih-simple-letter-agreement/](http://www.autm.net/resources-surveys/material-transfer-agreements/nih-simple-letter-agreement/). Accessed Dec. 12, 2015). General information is provided at [www.autm.net/autm-info/about-tech-transfer/about-technology-transfer/technology-transfer-resources/ubmta/](http://www.autm.net/autm-info/about-tech-transfer/about-technology-transfer/technology-transfer-resources/ubmta/). Accessed Dec. 12, 2015.

The Association of University Technology Managers (AUTM) holds the UBMTA Master Agreements from research institutions that wish to use the UBMTA for some or all of their exchanges of biological materials (Retrieved from [www.autm.net/resources-surveys/material-transfer-agreements/uniform-biological-material-transfer-agreement/](http://www.autm.net/resources-surveys/material-transfer-agreements/uniform-biological-material-transfer-agreement/). Accessed Dec. 12, 2015).

administrative burden for repositories and impose friction on commons-based sharing of mouse resources (Mishra and Bubela 2014).

Despite the existence of simplified mechanisms for materials exchanges, the use of complex, individually negotiated MTAs for research reagents is still common. Variable and negotiated MTAs rarely reflect the monetary value to the institution of the materials; indeed, they reflect a philosophical divide between institutionalized technology transfer professionals, tasked with institutional risk management and monetary returns on investment, and many researchers who wish to share their data and materials (Walsh et al. 2005a, 2005b; Walsh et al. 2007). In terms of risk management, research institutions are notoriously risk averse, but in reality, little litigation exists over exchanges for precompetitive research reagents compared to the volumes of MTAs negotiated each year (Bubela et al. 2015a). Indeed, our analysis of litigation found only 23 cases related to MTAs in the United States, of which only 4 concerned breaches of the terms of an MTA. Of interest to this chapter, although not directly relevant to mice, was an action brought by a biorepository – the American Type Culture Collection.<sup>5</sup> A researcher at the University of Pittsburgh was unsuccessfully prosecuted for mail fraud for using his institutionally approved account on behalf of a fellow researcher at an unapproved institution; the MTA prohibited transfer of materials to third parties. In our analysis, we do however recognize that more complex MTAs sometimes may be warranted, especially concerning exchanges of confidential information, when the contemplated transfers involve materials near clinical application or transfers to industry (Bubela et al. 2015a).

In the case of individually negotiated MTAs, AUTM and other innovation-focused institutions have further promulgated best practice guidelines that discourage some practices, such as reach-through terms that extend proprietary interests to derivative materials.<sup>6</sup> However, ironically, such terms were included in an effort to *promote* the research commons by the European component of the international knockout mouse project – EUCOMM. The clause in question entitled the Helmholtz Zentrum Munchen (the legal entity behind the EUCOMM Repository) “to a worldwide, nonexclusive, royalty-free, sublicensable and fully paid-up license to use, for noncommercial and teaching purposes, any IPRs that arise from the recipient’s use of the EUCOMM material” (Mishra and Bubela 2014). In other words, if the recipient developed a drug based on its research using EUCOMM material, then it was obligated to grant a license to the Helmholtz Zentrum Munchen on the terms specified in the EUCOMM MTA, which were broadly supportive of research commons. In our analysis, we concluded that such

<sup>5</sup> Order in *United States of America v. Steven Kurtz*, No. 04-CR-0155A (W.D.N.Y. Apr. 21, 2008), available at 2008 WL 1820903.

<sup>6</sup> Organisation for Economic Co-operation and Development. 2006. Guidelines for the Licensing of Genetic Inventions. Retrieved from [www.oecd.org/sti/biotech/guidelinesforthelicensingofgeneticinventions.htm](http://www.oecd.org/sti/biotech/guidelinesforthelicensingofgeneticinventions.htm); Association of University Technology Managers. 2007. In the Public Interest: Nine Points to Consider in Licensing University Technology. Retrieved from [www.autm.net/Nine\\_Points\\_to\\_Consider.htm](http://www.autm.net/Nine_Points_to_Consider.htm)

reach-through terms are equally problematic whether used to promote the nonprofit research or commercial interests because “the ability of repositories to monitor and then enforce this clause is questionable, and its complexity and presence may serve as a disincentive for potential users” (Mishra and Bubela 2014: 267).

The final complexity in materials sharing we wish to highlight is the transfer of materials to commercial entities, which raises the question of the extent to which private actors are part of the commons. In the mouse commons, commercial vendors, such as Charles River Laboratories, distribute research reagents to the pharmaceutical and biotechnology industries, as well as to nonprofit researchers. The biotechnology company Regeneron Pharmaceuticals Inc. was part of the Knockout Mouse Consortium because of its advanced capabilities in producing mouse models for pharmaceutical research.<sup>7</sup> Because of this commercial engagement in the development of the consortium in the United States, the KOMP repository does not restrict the distribution of its resources to the private sector. However, the situation was different in Europe, and here we tie back to the issue of background intellectual property rights over mouse-related materials and methods. The high-throughput resource generation centers needed to aggregate intellectual property rights over methods and processes used in their pipeline to construct the resource. Because of the complexity of the pipeline, however, it was not possible to identify all of the underlying IPRs and negotiate licenses for their use. The risk, therefore, was that the pipeline may have infringed IPRs, thereby making the resource-generating institutions vulnerable to patent infringement suits. It also limited the utility of the resource for industry use because in using a resource generated by infringing patents or incorporating patented materials, industry users faced patent infringement liabilities. Indeed, license negotiations over the underlying technologies continued until 2014 when a mechanism for distribution to industry was agreed upon by way of a French biotechnology company, genOway, which had aggregated identified underlying IP and took on the risk of distribution.<sup>8</sup>

The US members of the IKMC (the KOMP) were not so limited because of legal mechanisms available in that country (Bubela and Cook-Deegan 2015; Bubela et al. 2015b). In funding KOMP, the NIH employed a powerful legal tool commonly found in defense contracts – authorization and consent. The clause applies to patents and copyrights when use is “for and on behalf of the US

<sup>7</sup> Regeneron’s VelociMouse technology “enables the immediate generation of genetically altered mice directly from modified embryonic stem (ES) cells created by VelociGene, thereby avoiding the need to generate and breed chimeras (mice derived from a combination of modified and unmodified ES cells). This technology is faster and less expensive than other approaches currently being used. VelociMouse technology also enables the rapid creation of mice in which multiple modifications have been made in the same ES cells. This approach greatly reduces or eliminates extensive cross-breeding of mice that alters one gene at a time.” Retrieved from [www.regeneron.com/velocimouse](http://www.regeneron.com/velocimouse)

<sup>8</sup> [geoOway. 2016. Knockout and Reporter Mouse Catalogue. Retrieved from www.genoway.com/services/eucomm/eucomm-conditional-knockouts.htm?utm\\_source=google&utm\\_medium=cpc&utm\\_campaign=america](http://www.genoway.com/services/eucomm/eucomm-conditional-knockouts.htm?utm_source=google&utm_medium=cpc&utm_campaign=america)

government.”<sup>9</sup> When used in a research and development contract, its application means that the US government does not need to seek or negotiate a license to practice the patented invention. Moreover, 28 USC §1498 limits the government’s liability for patent infringement. While the patent holder is entitled to reasonable compensation, it cannot seek an injunction, damages, or lost profits against the government or a government contractor (in this case the members of KOMP) for patent infringement.<sup>10</sup> In effect, the US members of KOMP, unlike their European counterparts, were protected from potential patent infringement suits, enabling KOMP to distribute the mouse-related resources to industry. We argue in the next section that distribution to industry is essential for the sustainability of the research commons.

#### 10.5 GOVERNANCE FOR LONG-TERM SUSTAINABILITY OF THE COMMONS

The final issue we discuss here is the governance models needed to ensure that the mouse commons is sustainable over the long term, while remaining versatile and responsive to technological change (Mishra et al. 2016). The number and scale of repositories and databases supporting the mouse research commons to the global research community have expanded beyond their base, exemplified by JAX, that responded to need to share and distribute mice between individually funded research projects. The commons now includes archives that support high-throughput resource-generating initiatives. These initiatives, the IKMC and the IMPC, are international in scope, with member institutions in Austral-Asia, Asia, Europe, and North America and require sustainability of the mutant mouse resources developed by the IKMC and the IMPC (described earlier) beyond initial funding terms (Mishra et al. 2016). Here, we discuss three issues with respect to sustainability of the commons as it grows and undertakes additional tasks beyond sharing and distributing mouse models developed by individual research projects: (1) legal agreements that enable resources to be shared among and distributed by multiple archives within international consortia, (2) funding for archives, and (3) responsiveness to disruptive technologies.

The distribution of resources to researchers requires some mirroring/duplication of resources between repositories in different jurisdictions, both to ensure the security of the resource (e.g., against contamination or loss of funding for a repository) and to ease potential restrictions over the shipping of research materials to researchers across international borders. The sharing of resources across repositories within consortia remains problematic, however, because MTAs are required for such transactions (Mishra and Bubela 2014). Different drafting conventions and differences in ability to distribute to industry as opposed to only for noncommercial

<sup>9</sup> Judiciary and Judicial Procedure Act of 1948, 28 U.S.C. § 1498 (2016).

<sup>10</sup> The wording for and procedures relevant for granting authorization and consent are outlined in Federal Acquisition Regulations 52.227–1 and 27.2012–2, respectively.

research (discussed earlier) lead to difficulties in negotiating consortium agreements for the sharing of resources among repositories (Mishra and Bubela 2014). Long negotiations lead to delays, which have implications for the utility of the resources. Technological advances in this area of science are rapid, and new gene-editing technologies (Singh et al. 2015) may have superseded the utility of some aspects of the resources, in particular, the archive of mouse embryonic stem cells (Mishra et al. 2016).

The second challenge is the lack of sustained public funding. Funders commonly provide seed funding to establish archives, but such funding rarely provides a long-term investment to ensure sustainability of the resource (Mishra et al. 2016). Financial shortfalls, even if temporary, are problematic because they threaten the development and retention of highly qualified personnel and the maintenance of physical infrastructure and equipment. While there are economies of scale for larger archives, these also have higher operating costs, making them vulnerable to fluctuations in funding. Funders of archives generally demand a revenue-generation business model that transitions from outside funding to self-funding. Such models might incorporate differential pricing between noncommercial and commercial users, with the latter charged at a cost-recovery rate and the latter charged a premium rate that can be reinvested to support the operations of the archive. As discussed earlier, however, IP and other barriers to distribution to industry thus pose a problem for the financial sustainability of archives. In any event, given the realities of funding for the bulk of noncommercial users (and their decisions to allocate their research grants to purchasing research tools from archives), it is unlikely that archives will ever be self-sustaining; therefore, public funds will be needed to continue to subsidize archive operations (Mishra et al. 2016).

A further funding challenge is the lack of transnational funding models. In other words, archives operate transnationally, both in terms of depositors and distribution, but funding is national. In the context of archive consortia, national funders “find it difficult to harmonize policies and priorities to support archives in distinct jurisdictions but with networked operations” (Mishra et al. 2016: 284). Some European initiatives aim to develop new governance models to address the limitations of short-term national funding for research infrastructures, but such models are more difficult to implement across continents (Mishra et al. 2016). Thus the scale of transnational research commons limits their sustainability because of the lack of appropriate governance models that can facilitate long-term funding and prioritization of national funding agencies.

Finally, archives need to remain responsive to the utility of their resources and to technologies that may disrupt their business models. Mouse commons archives not only receive materials from small- and large-scale resource generators, store those materials, and distribute them, they also provide value-added, specialized services, such as in-depth phenotyping. Recent developments in gene-editing technologies, such as CRISPR (Clustered Regularly Interspersed Short Palindromic Repeats)

(Singh et al. 2015) are becoming increasingly cost effective and accessible to laboratories with expertise in molecular biology and use of animal models. Gene-editing technologies enable manipulation of genes in experimental organisms, such as mice, and allow for genomic alterations directly in embryos. They enable researchers to “avoid the lengthy and unpredictable process of genetically modifying [embryonic stem] cells, developing them into embryos and then into adult organisms, which may or may not be able to transmit the alterations to offspring” (Mishra et al. 2016: 287). This ability will likely reduce the reliance of researchers on archived mutant embryonic stem cell archives. Some repositories have already responded by providing gene-editing services to researchers who lack in-house capabilities, demonstrating the need for archives to adapt to new technologies to ensure long-term sustainability.

Ironically, gene-editing technologies may return the mouse-model research community to the conditions that the high-throughput resource generation projects and associated archives were designed to address, namely, non-sharing and non-standardization of mouse-related research reagents developed in individual research laboratories. As we explained in Mishra et al. (2016: 287–88):

Despite enthusiasm for genome-editing technologies, off-target effects (unwanted mutations elsewhere in the genome) are a serious issue and may make the reagents irreproducible [(Editorial 2014; Lin et al. 2014)] ... In the aggregate, the increased use of individualized “cottage industry” methods [namely gene-editing technologies] has several potential negative effects. It could divert already scarce research funds to the costly and piecemeal task of making reagents, which may have diminished quality and standardization in comparison with reagents produced using standard protocols in large resource-making efforts. Second, there are potential costs and losses from storing those reagents in ill-monitored and unstandardized small institutional freezers associated with contamination and damage.

In other words, these technologies may return the community to the “bad old days” when research groups were slow in depositing individualized reagents in archives because of the lack of incentives for doing so, or of consequences for not doing so. These conditions resulted in funder policies on data and materials sharing as well as funding for the high-throughput generation of standardized, high-quality community resources and support for associated archives. “Lessons learned are that widely dispersed resources lead to increased direct and marginal costs for funding agencies for the distribution of reagents associated with publications, duplication of resources, and increased mouse usage. The latter two effects counter the ethical experimentation aims of reduction, replacement and refinement, or ‘3 R’ (Russell and Burch 2014). Thus, these novel technologies may be disruptive not only to archives but also to norms of open, reproducible and ethical science” (Mishra et al. 2016: 288).

## CONCLUSION

In this chapter, we have discussed the lessons we have learned from the mouse model for human disease research community about rules and governance structures that sustain long-term access to international research infrastructures. With respect to rules, we focused on formal intellectual property rights and contracts in promoting or impeding research commons as well as rules development and implementation relevant to incentivizing, dis-incentivizing, and sanctioning participation in the commons. With respect to governance, we focused on policies that promote participation in the commons and funding and business models for long-term sustainability of the research commons.

The mouse research commons is international in scope and has a heterogeneous membership that varies in scale from small-scale laboratories, which develop individual mouse lines, to large-scale high-throughput generators of mouse-related research tools. This heterogeneity poses problems in facilitating the sharing of mouse-related research tools and an associated network effect, whereby resources are used and new knowledge and material are developed and re-contributed to the commons. Despite the development of archives that facilitate the sharing of mouse-related research tools through simplified rules for both sharing and relieving the logistical burdens of sharing for individual research laboratories, only 35 percent of new mice lines are shared. A plethora of policies that promote sharing are promulgated by funders and other research institutions; however, these are not accompanied by adequate enforcement strategies or incentive structures.

Contributing to the issue of incentives is the increased focus of funders on promoting the commercialization of publicly funded research outputs. Our analysis of the patent landscape indicated that the success of these incentives has contributed to the phenomenon of patenting of mouse-related research tools. While such patenting over compositions of matter has a small negative impact on follow-on research, the patenting of broad-based methods impacts the ability to develop high-throughput resource generation platforms that require the aggregation of multiple intellectual property rights. Protracted post hoc negotiations over such rights have delayed the distribution of resources, particularly to industry, a user-group that is essential to the financial sustainability of archives.

In conclusion, rules need to be put in place to incentivize use of and contributions to the commons. Where possible, MTAs and DTAs that mediate the exchange of materials and data, respectively, should be as simple as possible and avoid overly onerous terms that delay negotiations and the transfer of materials. This is especially problematic for low-value research tools in the precompetitive environment that are far from clinical application. Further governance structures are needed to address the international nature of the mouse commons. In reality, archives for research tools will require sustainable public funding to ensure their ongoing operations, utility, and ability to adapt to changing technologies and the needs of the user community. In our

opinion, the issues facing the mouse commons, and the solutions that have so far driven its evolution, are not unique. The advanced nature of the mouse commons in terms of rules-in-use and governance structures as well as the debates within the community related to these issues serve as models for other biomedical research commons that aim to support the translation of research from bench to bedside.

## REFERENCES

- Adams, D., R. Baldock, S. Bhattacharya, A. J. Copp, M. Dickinson, N. D. Greene, M. Henkelman, M. Justice, T. Mohun, S. A. Murray, E. Pauws, M. Raess, J. Rossant, T. Weaver, and D. West. 2013. Bloomsbury report on mouse embryo phenotyping: Recommendations from the IMPC workshop on embryonic lethal screening. *Disease Models & Mechanisms* 6 (3):571–79.
- Bradley, A., K. Anastassiadis, A. Ayadi, J. F. Battey, C. Bell, M.-C. Birling, J. Bottomley, S. D. Brown, A. Bürger, C. J. Bult, W. Bushell, F. S. Collins, C. Desaintes, B. Doe, A. Economides, J. T. Eppig, R. H. Finnell, C. Fletcher, M. Fray, D. Friendewey, R. H. Friedel, F. G. Grosveld, J. Hansen, Y. Héroult, G. Hicks, A. Hörlein, R. Houghton, M. Hrabé de Angelis, D. Huylebroeck, V. Iyer, P. J. Jong, J. A. Kadin, C. Kaloff, K. Kennedy, M. Koutsourakis, K. C. Kent Lloyd, S. Marschall, J. Mason, C. McKerlie, M. P. McLeod, H. Melchner, M. Moore, A. O. Mujica, A. Nagy, M. Nefedov, L. M. Nutter, G. Pavlovic, J. L. Peterson, J. Pollock, R. Ramirez-Solis, D. E. Rancourt, M. Raspa, J. E. Remacle, M. Ringwald, B. Rosen, N. Rosenthal, J. Rossant, P. R. Noppinger, E. Ryder, J. Z. Schick, F. Schnütgen, P. Schofield, C. Seisenberger, M. Selloum, E. M. Simpson, W. C. Skarnes, D. Smedley, W. L. Stanford, A. F. Stewart, K. Stone, K. Swan, H. Tadepally, L. Teboul, G. P. Tocchini-Valentini, D. Valenzuela, A. P. West, K.-I. Yamamura, Y. Yoshinaga, and W. Wurst. 2012. The mammalian gene function resource: The international knockout mouse consortium. *Mammalian Genome* 23 (9–10):580–86.
- Brown, S. D., and M. W. Moore. 2012a. The International Mouse Phenotyping Consortium: Past and future perspectives on mouse phenotyping. *Mammalian Genome* 23 (9–10):632–40.
- Brown, S. D., and M. W. Moore. 2012b. Towards an encyclopaedia of mammalian gene function: The International Mouse Phenotyping Consortium. *Disease Models & Mechanisms* 5 (3):289–92.
- Bubela, T., J. Guebert, and A. Mishra. 2015a. Use and misuse of material transfer agreements: Lessons in proportionality from research, repositories, and litigation. *PLoS Biology* 13 (2):e1002060.
- Bubela, T., S. Vishnubhakat, and R. Cook-Deegan. 2015b. The mouse that trolled: The long and tortuous history of a gene mutation patent that became an expensive impediment to alzheimer's research. *Journal of Law and the Biosciences* 23: 1–50.
- Bubela, T., P. N. Schofield, C. Ryan, R. M. Adam, and D. Einhorn. 2012a. Managing intellectual property to promote pre-competitive research: The



- mouse as a model for constructing a robust research commons. *Journal of Law and Information Science* 20 (2):98–121.
- Bubela, T., and R. Cook-Deegan. 2015. Keeping score, strengthening policy and fighting bad actors over access to research tools. *Nature Biotechnology* 33 (2):143–48.
- Bubela, T., G. A. FitzGerald, and E. R. Gold. 2012b. Recalibrating intellectual property rights to enhance translational research collaborations. *Science Translational Medicine* 4 (122):122cm3.
- Carbone, J., E. R. Gold, B. Sampat, S. Chandrasekharan, L. Knowles, M. Angrist, and R. Cook-Deegan. 2010. DNA patents and diagnostics: Not a pretty picture. *Nature Biotechnology* 28 (8):784–91.
- Caulfield, T., S. H. Harmon, and Y. Joly. 2012. Open science versus commercialization: A modern research conflict? *Genome Medicine* 4 (2):17.
- Collins, F. S., R. H. Finnell, J. Rossant, and W. Wurst. 2007. A new partner for the International Knockout Mouse Consortium. *Cell* 129 (2):235.
- Cook-Deegan, R., and C. Heaney. 2010a. Gene patents and licensing: Case studies prepared for the Secretary's Advisory Committee on Genetics, Health, and Society. *Genetics in Medicine* 12 (4 Suppl):S1–2.
- Cook-Deegan, R., and C. Heaney. 2010b. Patents in genomics and human genetics. *Annual Review of Genomics and Human Genetics* 11:383–425.
- Dedeurwaerdere, T. 2010a. Global microbial commons: Institutional challenges for the global exchange and distribution of microorganisms in the life sciences. *Research in Microbiology* 161:407–13.
- Dedeurwaerdere, T. 2010b. Self-governance and international regulation of the global microbial commons: Introduction to the special issue on the microbial commons. *International Journal of the Commons* 4:390–403.
- Editorial. 2014. Still much to learn about mice. *Nature* 509 (7501):399.
- Edwards, A. M., C. Bountra, D. J. Kerr, and T. M. Willson. 2009. Open access chemical and clinical probes to support drug discovery. *Nature Chemical Biology* 5 (7):436–40.
- Einhorn, D., and R. Heimes. 2009. Creating a mouse academic research commons. *Nature Biotechnology* 27 (10):890–91.
- Field, D., S. A. Sansone, A. Collis, T. Booth, P. Dukes, S. K. Gregurick, K. Kennedy, P. Kolar, E. Kolker, M. Maxon, S. Millard, A. M. Mugabushaka, N. Perrin, J. E. Remacle, K. Remington, P. Rocca-Serra, C. F. Taylor, M. Thorley, B. Tiwari, and J. Wilbanks. 2009. Megascience. 'Omics data sharing. *Science* 326 (5950):234–36.
- Fojo, T., S. Mailankody, and A. Lo. 2014. Unintended consequences of expensive cancer therapeutics – The pursuit of marginal indications and a me-too mentality that stifles innovation and creativity: The John Conley Lecture. *JAMA Otolaryngology Head Neck Surgery* 140 (12):1225–36.
- Friend, S. H. 2010. The need for precompetitive integrative bionetwork disease model building. *Clinical Pharmacology and Therapeutics* 87 (5):536–39.
- Frischmann, B. M., M. J. Madison, and K. J. Strandburg. 2014. Governing research commons. In *Governing Knowledge Commons*, B. M. Frischmann, M. J. Madison, and K. J. Strandburg eds., 1–44. Oxford: Oxford University Press.

- Grant, B. 2012. NIH to enforce OA policy. *The Scientist*. November 20. [www.the-scientist.com/?articles.view/articleNo/33327/title/NIH-To-Enforce-OA-Policy/](http://www.the-scientist.com/?articles.view/articleNo/33327/title/NIH-To-Enforce-OA-Policy/)
- Hay, M., D. W. Thomas, J. L. Craighead, C. Economides, and J. Rosenthal. 2014. Clinical development success rates for investigational drugs. *Nature Biotechnology* 32 (1):40–51.
- Heaney, C., J. Carbone, R. Gold, T. Bubela, C. M. Holman, A. Colaianni, T. Lewis, and R. Cook-Deegan. 2009. The perils of taking property too far. *Stanford Journal of Law, Science & Policy* 1:46–64.
- Hess, C., and E. Ostrom. 2006. A framework for analysing the microbiological commons. *International Social Science Journal* 188:335–49.
- Hess, C., and E. Ostrom. 2007. Introduction: An overview of the knowledge commons. In *Understanding Knowledge as a Commons: From Theory to Practice*, C. Hess and E. Ostrom eds., 3–26. Cambridge, MA: MIT Press.
- Hunter, J. 2014. Collaboration for innovation is the new mantra for the pharmaceutical industry. *Drug Discovery World* Spring:9–15.
- Jackson Laboratory. 2014. History of the Mouse Mutant Resource. Retrieved from [mousemutant.jax.org/MMR-history.html](http://mousemutant.jax.org/MMR-history.html).
- Kim, I. Y., J. H. Shin, and J. K. Seong. 2010. Mouse phenogenomics, toolbox for functional annotation of human genome. *BMB Reports* 43 (2):79–90.
- Kimmelman, J., and A. J. London. 2015. The structure of clinical translation: Efficiency, information, and ethics. *Hastings Center Report* 45 (2):27–39.
- Lin, Y., T. J. Cradick, M. T. Brown, H. Deshmukh, P. Ranjan, N. Sarode, B. M. Wile, P. M. Vertino, F. J. Stewart, and G. Bao. 2014. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research* 42 (11):7473–85.
- Mallon, A. M., V. Iyer, D. Melvin, H. Morgan, H. Parkinson, S. D. Brown, P. Flicek, and W. C. Skarnes. 2012. Accessing data from the International Mouse Phenotyping Consortium: State of the art and future plans. *Mammalian Genome* 23 (9–10):641–52.
- Mishra, A., P. N. Schofield, and T. M. Bubela. 2016. Sustaining large-scale infrastructure to promote pre-competitive biomedical research: Lessons from mouse genomics. *New Biotechnology* 33 (2):280–94.
- Mishra, A., and T. Bubela. 2014. Legal agreements and the governance of research commons: Lessons from materials sharing in mouse genomics. *OMICS* 18 (4):254–73.
- Munos, B. 2009. Lessons from 60 years of pharmaceutical innovation. *Nature Reviews Drug Discovery* 8 (12):959–68.
- Murray, F. 2010. The oncomouse that roared: Hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions. *American Journal of Sociology* 116 (2):341–88.
- Murray, F., P. Aghion, M. Dewatripont, J. Kolev, and S. Stern. 2009. Of mice and academics: Examining the effect of openness on innovation. *Working Paper No 14819*, National Bureau of Economic Research.
- National Center for Advancing Translational Sciences. 2015. Translational Science Spectrum. Retrieved from <https://ncats.nih.gov/translation/spectrum>

- Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action, Political Economy of Institutions and Decisions*. Cambridge: Cambridge University Press.
- Ostrom, E. 2005. *Understanding Institutional Diversity*. Princeton: Princeton University Press.
- Ostrom, E., and C. Hess. 2007. A framework for analyzing the knowledge commons. In *Understanding Knowledge as a Commons: From Theory to Practice*, C. Hess and E. Ostrom eds., 41–82. Cambridge, MA: MIT Press.
- Popp Berman, E. 2012. *Creating the Market University: How Academic Science Became an Economic Engine*. Princeton, NJ: Princeton University Press.
- Ringwald, M., and J. T. Eppig. 2011. Mouse mutants and phenotypes: Accessing information for the study of mammalian gene function. *Methods* 53 (4):405–10.
- Russell, W. M. S., and R. L. Burch. 2014. *The Principles of Humane Experimental Technique*. Maryland, MD: Johns Hopkins Bloomberg School of Public Health.
- Sauer, B., and N. Henderson. 1988. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proceedings of the National Academy of Sciences USA* 85 (14):5166–70.
- Schofield, P. N., T. Bubela, T. Weaver, L. Portilla, S. D. Brown, J. M. Hancock, D. Einhorn, G. Tocchini-Valentini, M. Hrabe de Angelis, and N. Rosenthal. 2009. Post-publication sharing of data and tools. *Nature* 461 (7261):171–73.
- Singh, P., J. C. Schimenti, and E. Bolcun-Filas. 2015. A mouse geneticist's practical guide to CRISPR applications. *Genetics* 199 (1):1–15.
- Skarnes, W. C., B. Rosen, A. P. West, M. Koutsourakis, W. Bushell, V. Iyer, A. O. Mujica, M. Thomas, J. Harrow, T. Cox, D. Jackson, J. Severin, P. Biggs, J. Fu, M. Nefedov, P. J. de Jong, A. F. Stewart, and A. Bradley. 2011. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* 474 (7351):337–42.
- Van Noorden, R. 2012. Key questions in the UK's shift to open-access research. *Nature News Blog*. Retrieved from <http://blogs.nature.com/news/2012/05/key-questions-in-the-uks-shift-to-open-access-research.html>
- Walsh, J. P., C. Cho, and W. M. Cohen. 2005a. Patents, material transfers and access to research inputs in biomedical research. *Final Report to the National Academy of Sciences' Committee Intellectual Property Rights in Genomic and Protein-Related Inventions*.
- Walsh, J. P., C. Cho, and W. M. Cohen. 2005b. View from the bench: Patents and material transfers. *Science* 309 (5743):2002–3.
- Walsh, J. P., W. M. Cohen, and C. Cho. 2007. Where excludability matters: Material versus intellectual property in academic biomedical research. *Research Policy* 36 (8):1184–203.
- Whitfield, J. 2011. Open access comes of age. *Nature* 474 (7352):428. doi: 10.1038/474428a.

APPENDIX: METHODS FOR MOUSE PATENT LANDSCAPE AND  
IMPACT ANALYSIS

We searched the Thomson database Delphion on September 27, 2007, for granted US patents using the following search strategy. First we used a modification of the Ade/Cook-Deegan algorithm.<sup>11</sup> The algorithm restricts the search to relevant patent classes and searches claims for terms commonly associated with DNA/RNA patents: ((((((119\* OR 426\* OR 435\* OR 514\* OR 536022\* OR 5360231 OR 536024\* OR 536025\* OR 800\*) <in> NC)

AND ((antisense OR <case><wildcard>cDNA\* OR centromere OR deoxyoligonucleotide OR deoxyribonucleic OR deoxyribonucleotide OR <case><wildcard>DNA\* OR exon OR “gene” OR “genes” OR genetic OR genome OR genomic OR genotype OR haplotype OR intron OR <case><wildcard>mtDNA\* OR nucleic OR nucleotide OR oligonucleotide OR oligodeoxynucleotide OR oligoribonucleotide OR plasmid OR polymorphism OR polynucleotide OR polyribonucleotide OR ribonucleotide OR ribonucleic OR “recombinant DNA” OR <case><wildcard>RNA\* OR <case><wildcard>mRNA\* OR <case><wildcard>rRNA\* OR <case><wildcard>siRNA\* OR <case><wildcard>snRNA\* OR <case><wildcard>tRNA\* OR ribonucleoprotein OR <case><wildcard>hnRNP\* OR <case><wildcard>snRNP\* OR <case><wildcard>SNP\*) <in> CLAIMS))

AND (((mouse) OR (mus\*) OR (mammal\*) OR (musculus) OR (murine) OR (mice) OR (Mus musculus))))

AND (((mammal\*) <in> CLAIMS) OR ((mouse) <in> CLAIMS) OR ((mus\*) <in> CLAIMS) OR ((murine) <in> CLAIMS) OR ((mice) <in> CLAIMS) OR ((musculus) <in> CLAIMS) OR ((Mus musculus) <in> CLAIMS)))

We then searched plant in claims (((Plant\*) <in> CLAIMS)) and removed all patents from search one that were also found in search two.

We downloaded all available data fields for the 7179 candidate granted patent identified by our search, including title, publication date, original national class, publication number, publication country, number of claims, assignee/applicant name, assignee/applicant state/city, assignee/applicant country, USPTO assignee code, USPTO assignee name, application number, application date, application country, attorney name, domestic references, number of domestic references, forward references, number of forward references, foreign references, other references, designated states national, designated states regional, ECLA codes, Examiner – primary, Examiner – assistant, family patent numbers, inventor name, inventor city/state, inventor country, IPC-R codes, inventive IPC-R, IPC-7 codes, Main IPC-7, National class, Main national class, field of search, maintenance status code, number of pages, priority number, priority date, and priority country.

<sup>11</sup> The algorithm for identifying DNA patents is described at <http://dnapatents.georgetown.edu/SearchAlgorithm-Delphion-20030512.htm>

## PATENT CODING

We read and coded all claims of all 7179 patents to (1) identify those patents that potentially claim mouse gene sequences; (2) identify the SEQ IDs of gene sequences actually claimed by patents; and (3) add additional codes, including: the assignee type (public/private university, government agency, pharmaceutical or biotechnology company, nongovernmental organization and individual inventor), any methods claimed, cell type(s) claimed, or transgenic animals claimed.

Included in our final analysis were 1144 patents that claimed mouse genes, mostly in the form of nucleotide sequences, but also amino acid sequences and a small number that claimed a gene by name. Prior to 1996, US patents did not require the genetic sequences to be listed with an associated SEQ ID.

## LIST OF PATENTED MOUSE GENE SEQUENCES

The resulting list of patent number–sequence ID pairs was matched, using a simple Python script written by postdoctoral fellow at the University of Alberta, Dr. Andreas Strotmann, against the Cambia Patent Lens database of genetic sequences extracted from US patents retrieved in June 2008.<sup>12</sup>

We retrieved, in FASTA format, nucleotide sequences for 32,351 DNA SEQ IDs in 929 patents and 179 amino acid SEQ IDs in 105 patents for a total of 32,530 sequences or sequence patterns listed in 983 patents (note that some patents listed both nucleotide and amino acid SEQ IDs). This data set was then manually filtered to retain only those sequences that were actually claimed in patents. We collected patented sequences that were not matched to the Patent Lens database from the Entrez database (if only the gene name was specified), from the patent claims themselves, or from the Patent Analysis website.<sup>13</sup>

## DETERMINING PATENTED MOUSE GENES

To determine the parts of the mouse genome that corresponded to the sequences in these patents, Dr. Songyan Liu, a bioinformatician, and his colleagues at the University of Manitoba performed a BLAST (basic local alignment search tool) analysis of all nucleotide and amino acid sequences identified earlier, using standard settings except for the following: Tag length  $\geq 25$ ; Expect  $< 0.001$ ; Score  $\geq 48$  (Figure 10.A1). The Expect value setting means that there is a less than 1 in 1000 chance that the gene match is the result of pure chance. This is significantly lower than in the usual bioinformatics setting but higher than the Expect=0 exact match

<sup>12</sup> Bacon, N., D. Ashton, R. A. Jefferson, and M. B. Connett. 2006. "Biological sequences named and claimed in US patents and patent applications, CAMBIA Patent Lens OS Initiative." Retrieved from [www.patentlens.net](http://www.patentlens.net)

<sup>13</sup> Patent Analysis. n.d. "Free databases." Retrieved from [www.patentanalysis.com/index.php?pagetype=news\\_databases](http://www.patentanalysis.com/index.php?pagetype=news_databases)

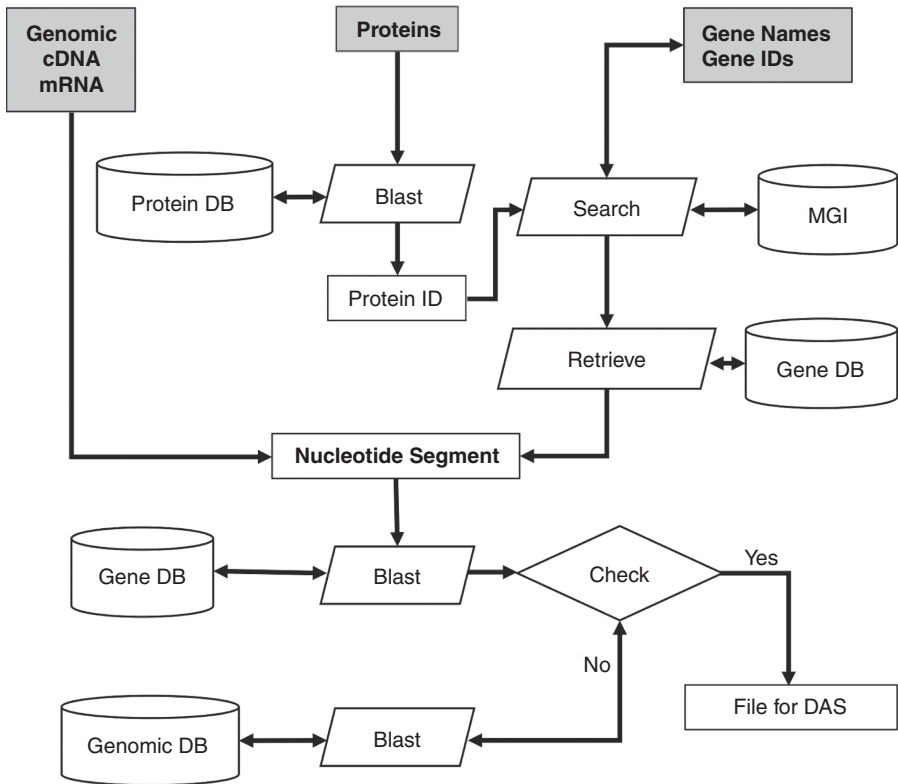


FIGURE 10.A1 Schematic of BLAST analysis – to match sequences identified from patents to the annotated mouse genome.

requirement in Murray and Jensen (2005).<sup>14</sup> The reasons for this choice are (1) most patent documents specifically state that they cover any genetic sequence *similar* to the one listed in the patent and (2) the sequence being patented and the corresponding sequence in the Ensembl database may be from different alleles of the same gene. In all cases, we retained only the best hit.

Using this method, we identified 1886 nucleotide sequences against the known mouse genome. An additional 62 entire genes were claimed by name or description rather than sequence. For the genes claimed by name or description, we searched the NCBI Entrez Gene database for entries matching their identifying description found in the patent claims. The resulting matches were added to the data set.

Our matching method identified 1692 genetic sequences from 952 mouse genes claimed, as a whole or in part, in 1049 US patent applications; including one mitochondrial gene (out of 37 known). This equates to 2.9% of the 32,480 mouse genes available in NCBI Mouse Build 37 against which we matched our sequences.

<sup>14</sup> Jensen, K., and F. Murray. 2005. "Intellectual property. Enhanced: Intellectual property landscape of the human genome." *Science* 310 (5746):239–40.

Other sequences were from unknown species with low homology to the mouse genome, were for noncoding fragments (i.e., did not map onto known mouse genes) or were artificial sequences.

#### COLLECTING INFORMATION ON PATENTED MOUSE GENES

For each of the 952 identified genes, Dr. Songyan Liu, at the University of Manitoba, and his colleagues extracted the following information from bioinformatics databases in December 2008:

- Trap hit: how many known hits were available for this gene.
- Gene targeting status: 822 of the patented genes (86%) had a corresponding targeting request at one of the knockout mouse consortia.
- OMIM information on the gene: 616 patented genes (65%) had an OMIM ID, 191 (20%) an OMIM description.<sup>15</sup>
- OMIM disease descriptors for 952 – (649 + 6) = 297 patented genes (31%).
- MGI phenotypes available for each gene: 485 of the patented genes had some kind of phenotype listed (51%).<sup>16</sup>
- Detailed Gene Ontology information per gene – all functions, processes, and components where this gene is known to play a role; 888 of the genes (93%) had one or more gene ontology entries.
- 945 genes (99.2%) had entries for all three gene ontology components in the MGI Gene Ontology Slim Chart;<sup>17</sup> that is, seven of the patented genes were still classified as “novel genes” at MGI at the time the searches were run.
- PubMed IDs for publications relevant to the gene.
- For mouse genes, this information is hand-curated by MGI and uploaded to the NCBI Entrez Gene database; 906 of the genes (95%) had corresponding PubMed publications.
- Human orthologues for the mouse gene: 866 of the genes (91%) had a known human orthologue.<sup>18</sup>
- MGI information: 883 of the genes (93%) had an MGI identifier.
- Coordinates for the gene’s position in the genome; this information is used for visualizations of the mouse gene patent landscape – it is available for all matched genes.

<sup>15</sup> McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD), and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD). 2009. “Online Mendelian Inheritance in Man, OMIM.”

<sup>16</sup> Bult, C. J., J. T. Eppig, J. A. Kadin, J. E. Richardson, J. A. Blake, and the members of the Mouse Genome Database Group. 2008. “The Mouse Genome Database (MGD): Mouse biology and model systems.” *Nucleic Acids Research* 36 (Database issue):D724–28.

<sup>17</sup> MGI. n.d. “MGI Gene Ontology GO\_Slim Chart Tool.” [www.informatics.jax.org/gotools/MGI\\_GO\\_Slim\\_Chart.html](http://www.informatics.jax.org/gotools/MGI_GO_Slim_Chart.html)

<sup>18</sup> MGI. 2016. “MGI reports: Human and mouse orthology.” Retrieved from <ftp://ftp.informatics.jax.org/pub/reports/index.html#orthology>

We also calculated statistics for genes using the MGI Gene Ontology Slim Chart Tool. These statistics were in addition to information specific to each genetic sequence mapped to each gene: Strand matched; Direction of match; Position of matched sequence in the genome; Chromosome (1-Y, mitochondrial); and Quality of match (score).

#### COMPARISON SET OF NON-PATENTED MOUSE GENES

For comparison purposes, Dr. Songyan Liu randomly selected a comparable number of unpatented genes. First, we randomly determined 2000 Ensembl Gene database entries for mouse genes. Of these, we removed 56 that were in the list of patented genes. Second, we searched for the remaining 1944 genes in MGI and identified 2012 hits. We removed 489 genes from this list if they did not have an official MGI symbol, 47 genes because they were in fact pseudogenes, and 96 genes because they were duplicates, including genes with multiple loci or Y chromosome genes that were a duplicate of X chromosome genes. In total, therefore, we selected 1397 genes for the control set to compare against our 952 patented genes out of a total of 32,480 possible genes (including mitochondrial genes) from NCBI Mouse Build 37.

As earlier, we extracted the following information in February 2009 on the genes from bioinformatics databases:

- 1069 genes (77%) had been investigated for targeting.
- All genes in this control set had hits in all three components of the MGI Gene Ontology Slim Chart (i.e., none were “novel genes”).
- 144 genes (10%) had a corresponding OMIM ID; 133 (9.5%) had associated detailed disease identifiers.
- 266 genes (19%) had associated phenotype information.
- 1079 (77%) had at least one component of associated Gene Ontology information.
- 1211 (87%) had associated PubMed publications.

#### MOUSE GENE LITERATURE

We downloaded the full XML records for the MGI mouse gene associated PMIDs from PubMed, which resulted in 23,805 publications on patented mouse genes and 10,684 on non-patented mouse genes in December 2008. We then downloaded full records for literature that cited those publications from Thomson’s ISI database. In detail, we

- Parsed XML PubMed records into an SQL database, using a Python script written by Dr. Strotmann, to extract (1) author names, affiliation; (2) article title, major MeSH codes; and (3) journal name, issue, year, number, pages.



- Located and downloaded full corresponding records in the Thomson ISI database so that we could download all citing literature. We located 98% of PubMed records in ISI.

### STATISTICAL ANALYSIS

The goal of our statistical analysis, performed by consulting biostatistician Dr. Shawn Morrison, was to determine if the citation and publication rates for publications on mouse genes (1) changed after patenting and (2) differed between publications on patented and unpatented mouse genes. We considered eight time periods:  $\pm 1$ ,  $\pm 3$ ,  $\pm 5$ , and  $\pm 7$  years before and after patenting. For patented genes, date “o” was the date the US patent was granted. For non-patented genes, date “o” was the median time from the original publication to the date of the search. This gave us a distribution of publications that had from o to at least 14 years of publication and citation data. Given the length of time from scientific publication to patent grant, the two data sets had similar distributions around the patent date and the median publication date.

We retained only those articles that had sufficient data to estimate all year intervals for analysis. For example, if an article had  $\pm 4$  years of data, it was included in the  $\pm 3$  years analysis, but not the  $\pm 5$  analysis. Some genes had sufficient data for the pre-patenting period but not the post-patenting period (and vice versa), and therefore sample sizes vary for each period.

Data in the original data set was on a per article basis (citations per year and per article). We re-summarized this information on a *per gene* basis rather than a *per article* basis. For example, in a given year, if one article about gene ‘X’ was cited 10 times, and another article about gene ‘X’ was cited 5 times, then the result was a total of 15 citations for that gene in that year. This per gene data was used to calculate citation rates and was the basis for summary statistics and t-tests (described later).

We calculated the publication and citation rates per gene for the eight periods. Calculation of citation rate requires information regarding the change in the number of publications/citations from one year to the next. For example, the citation rate in the first year post-patenting would be the rate from Year o to Year 1, the rate for the second year would be the rate from Year 1 to Year 2, and so on. More formally, the citation rate was the natural log of the ratio between the years of interest – this provides an estimate of the instantaneous rate of change at that point in time (i.e., the slope).

Some genes had a number of publications/citations in a given year but declined to zero citations in the next. This created difficulties in calculating rates (i.e., division by zero), and these genes were excluded from analysis. Fortunately, this only applied to a relatively small number of genes. The exception to this filtering rule occurs when both the starting and ending years had zero citations. In this case, the rate was unchanged (and calculated as a rate of change = 0.00).

TABLE 10.A1 *Rate calculations for publications*

Period of Interest (relative to patent year)	Year of Citation Data per Gene		Rate Calculation
	From	To	
-1	0	-1	$\ln(\text{pubs in year-1/ pubs in year 0})$
+1	0	+1	$\ln(\text{pubs in year 0/ pubs in year+1})$
-3	-3	-2	$\ln(\text{pubs in year-3/ pubs in year -2})$
+3	+2	+3	$\ln(\text{pubs in year+2/ pubs in year+3})$
-5	-5	-4	$\ln(\text{pubs in year-5/ pubs in year-4})$
+5	+4	+5	$\ln(\text{pubs in year+4/ pubs in year+5})$
-7	-6	-7	$\ln(\text{pubs in year-7/ pubs in year-6})$
+7	+6	+7	$\ln(\text{pubs in year+6/ pubs in year+7})$

Therefore, the years used in the calculation of publication rate for this analysis are shown in Table 10.A1 (note that the same rate calculation was applied to citations).

#### SAMPLE CALCULATIONS AND CONVERSIONS

If an article was cited 10 times in the year of patent grant (Year 0) and cited 11 times in the year following (Year 1), then the rate of citation during the first year post-patenting (Year 0 to Year 1) would be:

$$\text{Citation Rate} = \ln \left[ \frac{11}{10} \right] = 0.09531$$

To estimate the percentage increase in citations over a given period, it is necessary to convert the instantaneous rate of change ( $r$ ) to the finite rate of change ( $\lambda$ ) as follows:

$\lambda = e^r$ , where “ $\lambda$ ” is the finite rate of change and “ $r$ ” is the instantaneous rate of change.  $\lambda$  may be thought of as a “multiplier” between years. In the previous example, one would have to have an increase of 10% for the number of citations to increase from 10 to 11. The multiplier in this situation is 1.1, or a 10% increase.

For example, if  $r = 0.09531$ , then the finite citation rate is calculated as  $e^r = e^{0.09531} = 1.1$  per year, which is interpreted as a 10% increase in the number of citations. To convert back, the equation is as follows:  $\ln(\lambda) = r = \ln(1.1) = 0.09531$ .

The relationship between  $r$  and  $\lambda$  is shown in the accompanying table.

Citation Rate Is	$r$	$\lambda$
Decreasing	$<0.0$	$<1.0$
Stable	$=0.0$	$=1.0$
Increasing	$>0.0$	$>1.0$

Thus, the citation rate is increasing when  $r > 0$  and/or  $\lambda > 1.0$ .

#### ANALYSIS AND RESULTS

Summary statistics for publication and citation rate per gene were calculated for each time period (Tables 10.A2–10.A5). Time periods were compared using Welch's t-tests<sup>19</sup> which are similar to the common Student's t-test but without the requirements for equal variances or equal sample sizes. A t-test was conducted for each period ( $\pm 1$ ,  $\pm 3$ ,  $\pm 5$  and  $\pm 7$  years pre- and post-patenting) within publications and citations. Welch's t-tests were then used to compare each time period between patented and unpatented genes for both publication and citation rates. To compensate for false positive significance as a result of large sample sizes and multiple t-tests, we increased the significant P-value from 0.05 to 0.01. In the following tables, significant differences are bolded. In addition for each time period, we compared publication and citation rates between patented and unpatented genes using Welch's t-test (Table 10.A6).

TABLE 10.A2 Summary statistics for publication rate per patented gene

Period Relative to Patent Grant Year	Mean (r)	Std. Error	# genes	# publications	Years Compared	P-value	t-value	df
1 year prior	-0.009	0.020	633	2084	-1 to +1	0.091	1.691	4159
1 year post	-0.059	0.021	606	2088				
3 years prior	-0.008	0.019	619	2332	<b>-3 to +3</b>	<b>0.007</b>	<b>2.710</b>	<b>4043</b>
3 years post	-0.082	0.019	633	1767				
5 years prior	-0.034	0.021	649	2279	-5 to +5	0.873	0.160	3504
5 years post	-0.030	0.014	696	1240				
7 years prior	0.027	0.011	706	1890	-7 to +7	0.564	0.578	2386
7 years post	-0.027	0.011	745	757				

<sup>19</sup> Welch, B. L. 1947. "The generalization of 'student's' problem when several different population variances are involved." *Biometrika* 34:28–35.

TABLE 10.A3 Summary statistics for citation rate per patented gene.

Period Relative to Patent Grant Year	Mean (r)	Std. Error	# genes	# citations	Years Compared	P-value	t-value	df
1 year prior	0.417	0.028	680	118037	-1 to +1	0.052	1.94	237635
1 year post	0.344	0.024	738	144949				
3 years prior	0.473	0.032	490	73537	-3 to +3	<.0001	7.276	157969
3 years post	0.179	0.024	779	193711				
5 years prior	0.568	0.043	325	39525	-5 to +5	<.0001	7.965	66863
5 years post	0.179	0.024	740	222987				
7 years prior	0.572	0.063	184	21173	-7 to +7	<.0001	8.806	28759
7 years post	-0.030	0.026	647	236608				

TABLE 10.A4 Summary statistics for publication rate per unpatented gene.

Period Relative to Median Publication Date	Mean (r)	Std. Error	# genes	# publications	Years Compared	P-value	t-value	df
1 year prior	-0.026	0.020	556	1025	-1 to +1	0.668	0.429	1988
1 year post	-0.037	0.015	798	1444				
3 years prior	-0.018	0.015	700	1432	-3 to +3	0.770	0.292	2126
3 years post	-0.013	0.010	875	702				
5 years prior	0.0003	0.012	912	1063	-5 to +5	0.624	0.491	1280
5 years post	-0.006	0.005	1080	220				
7 years prior	0.029	0.008	1050	894	-7 to +7	0.038	2.086	261
7 years post	-0.001	0.004	1125	122				

TABLE 10.A5 Summary statistics for citation rate per unpatented gene.

Period Relative to Median Publication Date	Mean (r)	Std. Error	# genes	# citations	Years Compared	P-value	t-value	df
1 year prior	0.632	0.027	960	44835	-1 to +1	<.0001	15.801	98752
1 year post	0.606	0.023	1027	73640				
3 years prior	0.602	0.046	332	14329	-3 to +3	<.0001	11.593	17770
3 years post	0.040	0.016	1187	171824				
5 years prior	0.332	0.051	154	8652	-5 to +5	<.0001	4.047	10459
5 years post	0.116	0.016	1142	224631				
7 years prior	0.204	0.059	109	5287	-7 to +7	<.0001	16.046	7113
7 years post	-0.812	0.024	1119	235114				

TABLE 10.A6 *Summary statistics for comparison within each time period of publication rate for patented and unpatented genes*  
*Statistically significant differences are bolded.*

Period Relative to Year 0*	P-value	t-value	df
7 years prior	0.896	0.130	2731
5 years prior	0.147	1.451	3260
3 years prior	0.679	0.414	3760
1 year prior	0.560	0.584	2747
1 year post	0.389	0.862	3418
3 years post	<b>0.001</b>	<b>3.211</b>	<b>2407</b>
5 years post	0.114	1.583	1449
7 years post	0.023	2.284	873

TABLE 10.A7 *Summary statistics for comparison within each time period of citation rate for patented and unpatented genes*  
*Statistically significant differences are bolded.*

Period Relative to Year 0*	P-value	t-value	df
7 years prior	<.0001	<b>4.266</b>	<b>18451</b>
5 years prior	<b>0.0004</b>	<b>3.547</b>	<b>22556</b>
3 years prior	0.022	2.294	30776
1 year prior	<.0001	<b>5.467</b>	<b>133264</b>
1 year post	<.0001	10.570	204122
3 years post	<.0001	<b>4.851</b>	<b>325261</b>
5 years post	0.028	2.203	392287
7 years post	<.0001	<b>22.378</b>	<b>467887</b>

\* Note that in Tables 10.A6 and 10.A7, year 0 is the year the patent was granted for patented genes and the median year for publications from first publication to date of search for unpatented genes (December 2008) for unpatented genes.