Week 2: Intellectual Property

Carolyn Stein

Econ 220C: Topics in Industrial Organization

Annoucements

- Problem set 1 has been posted
- Ending class at 11:30 today

Does stronger IP lead to more innovation?

- Moscona (2022)
- Budish, Roin, and Williams (2015)

Does IP inhibit follow-on innovation?

- Williams (2013)

Does disclosure matter?

- Sampat and Williams (2019)

Why not just ask firms?

Mansfield (1986) randomly surveyed 100 firms from 12 industries. Asked firms "what share of innovations would not have been developed / commercially produced in the absence of patent protection?"

Table 1 Patents and research investments: evidence from the Mansfield (1986) survey

Industry	Percent that would not have been developed	Percent that would not have been introduced
Pharmaceuticals	65	60
Chemicals	30	38
Petroleum	18	25
Machinery	15	17
Fabricated metal products	12	12
Primary metals	8	1
Electrical equipment	4	11
Instruments	1	1
Office equipment	0	0
Motor vehicles	0	0
Rubber	0	0
Textiles	0	0

Percent of developed or commercially introduced inventions that would not have been developed or commercially introduced if patent protection could not have been obtained, 12 industries, 1981–1983. Data taken from Mansfield (1986, table 1).

Are these the industries you'd expect to be most responsive?

Survey evidence $\neq$ evidence?

Boldrin and Levine write in "The Case Against Patents" (JEP, 2013)

- "There is no empirical evidence that [patents] serve to increase innovation and productivity, unless productivity is identified with the number of patents awarded – which evidence shows, has no correlation with measured productivity"
- However, it is difficult to find the right variation!
- Lerner (2009) studies 177 changes in patent policy in 60 countries and finds no innovation response (in either the focal country or other countries)
- However innovations are often developed for a global market and so changes in one small country may not be very meaningful

Does stronger IP lead to more innovation?

- Moscona (2022)
- Budish, Roin, and Williams (2015)

Does IP inhibit follow-on innovation?

- Williams (2013)

Does disclosure matter?

- Sampat and Williams (2019)

Direct evidence from agricultural biotechnology

Moscona (2022) provides direct evidence from seed development. Institutional details make this possible:

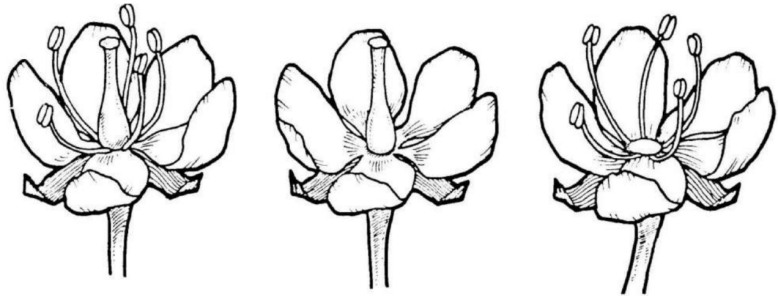
- Seeds were excluded from patent protection until 1985 with the *Ex parte Hibberd* decision
  - "Virtually overnight, and to the great surprise of many, seeds became patentable" – William Lesser (1987)
- This meant that prior to 1985, farmers could save and re-use seeds, making it harder for innovators to profit
- However, some types of plants had *de facto* protection prior to this ruling $\rightarrow$ should not have been impacted (and can serve as a control group!)

Hybrid vs. non-hybrid seeds

- One way of creating innovative seeds is hybridization
  - Cross a plant (à la Mendel) with genes AA with a second plant with genes aa to get a child plant with genes Aa
  - But what about the children of this plant? Only 50% will have the desired Aa genes
  - This problem gets worse the more genes you care about
  - Key idea: *farmers cannot harvest and reuse the seeds* $\rightarrow$ implicit IP protection!
- Other ways of innovating:
  - Paper is vague on this!
  - Genetic modification (but this starts post-1985)

A short plant biology lesson: perfect vs. imperfect flowers

Plant flowers come in one of two varieties: perfect or imperfect



(a) Perfect Flower (b) Imperfect Flower (Pistillate) (c) Imperfect Flower (Staminate)

Figure 1: **Perfect vs. Imperfect Flowers.** This figure shows the distinction between perfect flowers, which contain both reproductive organs (Figure 1a), and imperfect flowers, which have either only the "female" (Figure 1b) or only the "male" (Figure 1c) reproductive organ.

- Perfect flowers have both male and female reproductive organs in the *same flower*
- Imperfect flowers have these in *different flowers*

Why does this matter?

- **Imperfect flowers** are easy to hybridize – you can manually move pollen from the stamen of one flower to the pistil of another. Therefore plants with imperfect flowers are the *control group*: they already had built-in IP protection
- **Perfect flowers** are very difficult to hybridize – they will self-pollinate. Moreover, this self-pollination means that seeds will be consistent from one generation to the next. Farmers can re-use the seeds $\rightarrow$ the benefits of patenting are large. Therefore plants with perfect flowers are the *treatment group*

These data were collected by the author by hand from over 300 sources

Validating the hybridization claim

Imperfect flowers are 60% more likely to have hybrid seed varieties

Table A1: Imperfect Flowers and Hybrid Technology in the Patent Data

	(1)	(2)
	Word "Hybrid" in Title or Abstract (=1)	
Sample includes all patents in CPC class:	A01H	A01H.5 or A01H.6
Crop with Imperfect Flowers (=1)	0.597***	0.622***
	(0.0169)	(0.0171)
F-statistic	1247.13	1332.38
Month Fixed Effects	Yes	Yes
Observations	7,096	6,685
R-squared	0.268	0.284

Notes: The unit of observation is a patent and the sample includes all patents applied for prior to 2000 in CPC class A01H (column 1) or CPC classes A01H.5 or A01H.6 (column 2). All specifications include month fixed effects and robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Measuring innovation

- Patent counts will not work – since seeds were not eligible for patent protection pre-1985, we would see no seed patents even if there was lots of innovation!
- Number of new crops: USDA *Variety Name List* – collects data on all released varieties to prevent fraud in the seed market
- Investment: project-level R&D data from the USDA Current Research Information System. Funding is categorized by crop. Also flags if research received public funding

Empirical strategy: difference-in-differences

Main specification for crop c in year t:

$$y_{ct} = \alpha_c + \delta_t + \beta \cdot \text{Not Hybrid}_c \cdot \mathbb{I}_t^{\text{Post 1985}} + X_{ct}'\Gamma + \varepsilon_{ct}$$

where Not Hybrid is an indicator for whether the crop has perfect flowers

β is the coefficient of interest, as it shows the difference in trend between the treatment (perfect /non-hybrid) and control (imperfect / hybrid) plants

Static diff-in-diff results

Column (1) implies that treated crops had $e^{0.795} - 1 \approx 120\%$ more varieties following the introduction of patent rights

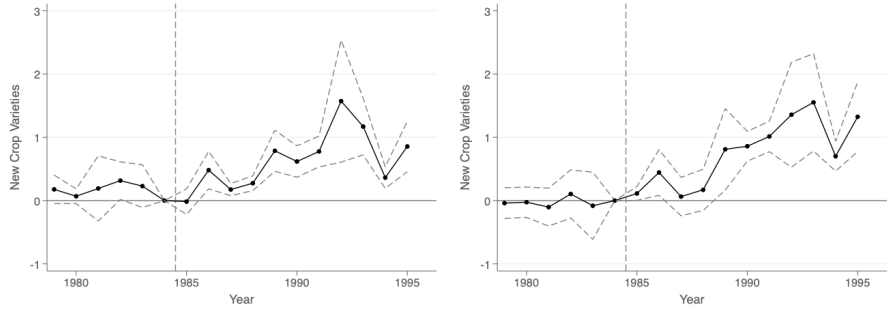
Table 2: Patent Protection and Novel Plant Varieties

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	New Varieties (count)				New Varieties (asinh)			
Specification:	Poisson	Poisson	Poisson	Poisson	OLS	OLS	OLS	OLS
Not Hybrid Compatible$_i$ x $\mathbb{1}_t^{\text{Post 1985}}$	0.795***	0.920***	0.931***	0.919***	0.109***	0.204**	0.275**	0.320***
	(0.240)	(0.322)	(0.327)	(0.329)	(0.0334)	(0.0837)	(0.111)	(0.105)
Crop Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GMO Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Rain Sensitivity Controls	No	No	Yes	Yes	No	No	Yes	Yes
Reproduction Type Controls	No	No	No	Yes	No	No	No	Yes
Observations	2,260	2,260	2,060	2,060	2,280	2,280	2,060	2,060
R-squared	-	-	-	-	0.815	0.819	0.826	0.829

Notes: The unit of observation is a crop-year. All specifications include crop and year fixed effects. In columns 1-4, the outcome variable is the number of new varieties and in columns 5-8, it is the inverse hyperbolic sine transformation of the number of new varieties. The regression model is noted at the top of each column. The controls are listed at the bottom of each column and are included as a fixed crop-level characteristic interacted with a full set of year fixed effects. Standard errors, double clustered by crop and year, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Dynamic diff-in-diff results

Pre-trends in β look flat:



(a) No Controls

(b) All Controls

Figure 4: **Patent Protection and Novel Plant Varieties Over Time.** Coefficient estimates from poisson estimates of Equation (3) The dependent variable is the number of novel plant varieties in the crop-year. Standard errors are double clustered by crop and year; 95% confidence intervals are reported.

Supporting evidence for the profit story

- Perennial crops live for 2+ years; effects are larger for non-perennial crops
- Effects on R&D spending are driven by private rather than public spending

(a) Public Research Investment

(b) Private Research Investment

Figure 5: **Research Investment.** Coefficient estimates from Equation (3). The dependent variables are the inverse hyperbolic sine of public research investment (a) or private research investment (b) in the crop-year pair. All baseline controls are included in each specification. Standard errors are double-clustered by crop and year; 95% confidence intervals are reported.

Evidence of downstream effects

- Crop yields increase for treated crops
- Counties most exposed to treated crops see increases in land values (which author argues are due to increased farm profits).
- Suggests that innovation benefit more than offsets the monopoly costs

Does stronger IP lead to more innovation?

- Moscona (2022)
- Budish, Roin, and Williams (2015)

Does IP inhibit follow-on innovation?

- Williams (2013)

Does disclosure matter?

- Sampat and Williams (2019)

What about in the pharmaceutical industry?

- Mansfield survey suggests this might be the most important field
- But no obvious sources of exogenous variation in patent protection
- Budish, Roin, and Williams (2015) show that you can still make progress...

Canonical example: lung cancer

- From 2010-2015, eight drugs have been approved to treat lung cancer
- All eight of these drugs treat *late stage* lung cancer and only yielded incremental improvements in survival
  - Example: Genentech's drug Avastin extends life from 10.3 months to 12.3 months on average
- Why the focus on late-stage cancer?
  - Maybe just scientifically easier to develop?
  - But also can be brought to market faster...

Time to market and effective patent life

- Plenty of theory (and some evidence) that firms are excessively short-term focused
- However, this problem may be particularly acute in pharmaceutical due to a quirk of the patent system:
  - Fixed 20-year patent term starts at the time of invention, rather than the time of commercialization
  - For innovations where the clinical trial takes a long time, the *effective* patent life can by much shorter than 20 years!
  - Can create a strong incentive to develop drugs that will go through clinical trial quickly

Cancer stage is predictive of commercialization time

- FDA requires that a drug shows efficacy in clinical trial
- For cancer drugs, increased efficacy means improved survival
- Later stage cancer $\rightarrow$ shorter survival $\rightarrow$ faster to show efficacy
- Cancer setting / data has two important features:
  - By gridding the cancer space into disease x stage cells, you can see where innovation is happening and where it is *missing*
  - Even for the missing innovation, the authors can compute commercialization time using five-year survival rate

Descriptive evidence

Far more R&D occurs for late-stage cancers:

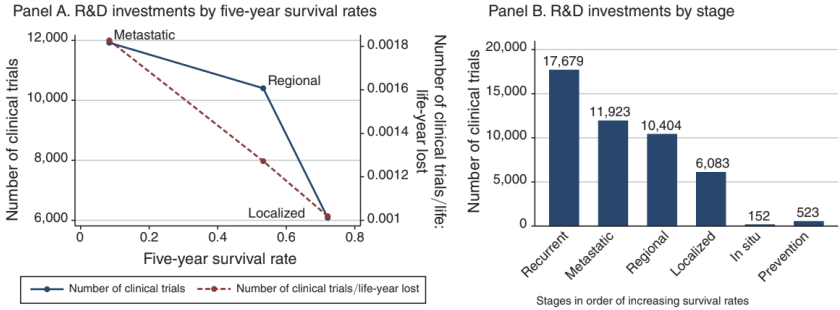


FIGURE 1. SURVIVAL TIME AND R&D INVESTMENTS: STAGE-LEVEL DATA

More formally...

Use patient-level data from the Surveillance, Epidemiology, and End Results (SEER) database to compute disease x stage 5-year survival rates:

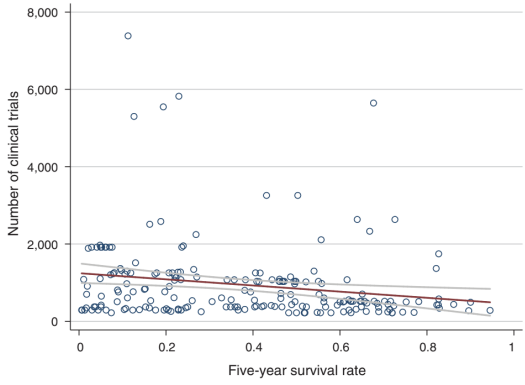


FIGURE 2. SURVIVAL TIME AND R&D INVESTMENTS: CANCER-STAGE DATA

- Relationship is robust to controls for market size / years of life lost
- But what if late-stage cancers just have more R&D opportunities?

Supporting evidence I: surrogate endpoints

- If a short-run outcome is very good a predicting a long-run outcome, you can evaluate the short-run outcome instead. This is called a *surrogate outcome*
- In this context: is there a good way of predicting survival without having to wait and see if patients survive?
- For leukemias (blood cancers): yes!
  - White blood cell counts are established as a non-mortality endpoint
  - This has long been understood: When Sidney Farber discovered chemotherapy in 1948, he initially focused on leukemias precisely because he knew he could monitor white blood cell counts to assess progress

Surrogate endpoints break relationship between 5-year survival and R&D

This is inconsistent with late-stage cancer R&D being less scientifically feasible:

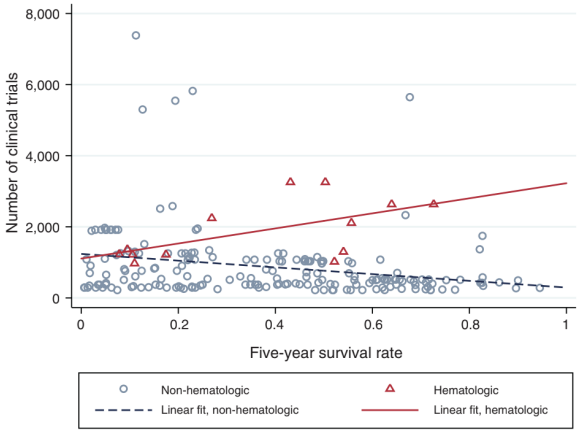


FIGURE 4. SURROGATE ENDPOINTS, SURVIVAL TIME, AND R&D INVESTMENTS

Supporting evidence II: publicly vs. privately funded trials

If the lack of trials for slow-to-commercialize cancers is due to scientific infeasibility, this should affect publicly and privately funded trials equally. But this is not the case:

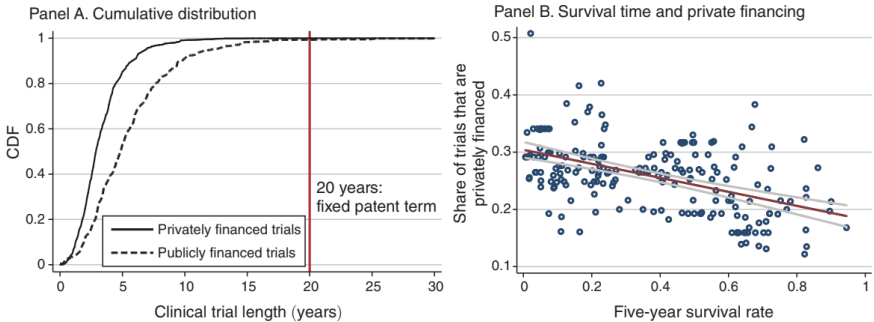


FIGURE 5. SURVIVAL TIME AND FINANCING OF CLINICAL TRIALS

Computing the missing R&D

Compare 5-year survival rates in 2003 to 1973 to see where progress has happened and where it is "missing." Valuing human life years at $100,000 per year implies a cost of $89 billion per patient cohort from this missing R&D

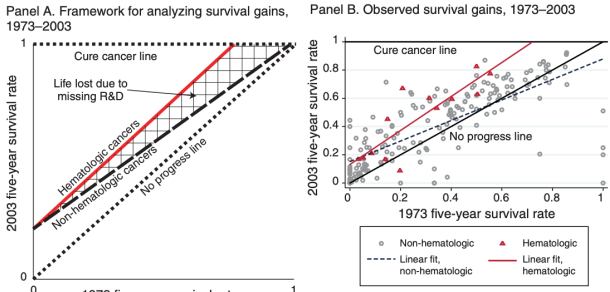


FIGURE 6. SURVIVAL GAINS, 1973–2003

Notes: This figure illustrates how we use variation in surrogate endpoints (across hematologic and non-hematologic cancers) to estimate counterfactual survival gains from 1973–2003 that would have been observed had commercialization lags for non-hematologic cancers mirrored the shorter commercialization lags realized for hematologic cancers. Panel A illustrates our conceptual framework. Panel B illustrates the empirical analog of panel A, plotting the 1973 five-year survival rate against the 2003 five-year survival rate. The level of observation is the cancer-stage. For details on the sample, see the text and online Data Appendix.

Takeaways

- Amazing how much you can do with no exogenous variation!
  - If you use theory to structure your thinking
  - And understand the setting well enough to come up with other empirical tests
- Worth noting that the authors cannot decompose how much of the bias against long t_{comm} drugs is due to corporate short-termism ($\eta < 1$) vs. the fixed patent term: the model makes this clear. There are no predictions that separate the two
- Surrogate outcomes are a clear policy winner! And we should perhaps invest in developing more of them

Does stronger IP lead to more innovation?

- Moscona (2022)
- Budish, Roin, and Williams (2015)

Does IP inhibit follow-on innovation?

- Williams (2013)

Does disclosure matter?

- Sampat and Williams (2019)

Intellectual property rights and follow-on innovation

- Suppose a Genetic Sequencing Corp holds IP on a gene
- Suppose that Pfizer discovers a genetic diagnostic based on the private firm's gene
- Will this IP discourage Pfizer from developing the test?
- Challenge: genes with IP may not be randomly selected...

The Human Genome Project

- The Human Genome (HGP) Project began in 1990
- Coordinated public effort to sequence the entire human genome and put all genes *in the public domain*
- In 1999, the private firm Celera also entered the race
- Between 2001-2003, Celera used a form of IP to protect the genes it sequenced that the HGP had not yet sequenced
  - This IP meant that firms had to negotiate licensing agreements with Celera if they made a commercial discovery that used the gene
  - Note that this IP was not actually a patent

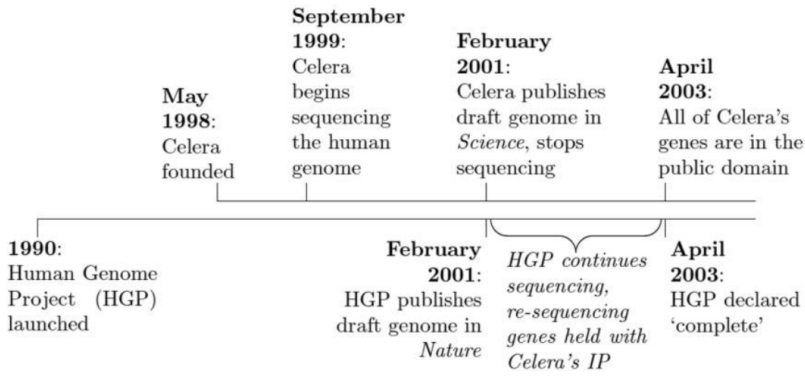
Data construction

The unit of analysis is a single gene. Consider the gene RAX2:

- The gene sequence appears one time in RefSeq in 2001
- The gene was never held be Celera's IP
- From the Online Mendelian Inheritance in Man (OMIM) database, we can see the gene is linked to **two** phenotypes, both of which reference **one** publication:
  - Publication is from *Human Molecular Genetics* (2004)
  - Phenotype 1 is age-related macular degeneration
  - Phenotype 2 is cone-rod dystrophy
- GeneTests.org lists **one** genetic test for RAX2 for age-related macular degeneration

Timeline

Key events:



Simple tabulation suggests that follow-on innovation suffered

TABLE 1
INNOVATION OUTCOMES FOR CELERA AND NON-CELERA GENES SEQUENCED IN 2001

	Celera Mean (1)	Non-Celera Mean (2)	Difference [(1) − (2)] (3)	*p*-Value of Difference (4)
Publications in 2001–9	1.239	2.116	−.877	[.000]
1(known, uncertain phenotype)	.401	.563	−.162	[.000]
1(known, certain phenotype)	.046	.073	−.027	[.000]
1(used in any diagnostic test)	.030	.054	−.024	[.000]
Observations	1,682	2,851		

NOTE.—This table compares subsequent innovation outcomes for Celera genes relative to non-Celera genes sequenced in the same year. Gene-level observations. The sample in col. 1 includes all Celera genes; the sample in col. 2 includes all non-Celera genes sequenced in 2001. The *p*-value reported in col. 4 is from a *t*-test for a difference in mean outcomes across cols. 1 and 2. See the text and online App. A for more detailed data and variable descriptions.

But maybe Celera's genes were just "worse"?

Concerns about selection

- HGP intentionally sequenced the most scientifically promising genes first
- Celera used a method called “shotgun sequencing” that gave them little control over which genes they sequenced
- Combination implies that negative selection into Celera IP is a serious concern

Testing for selection

- One way to measure ex-ante value of a gene is the number of papers written about it prior to it being sequenced (this is available in OMIM)
- Suggests Celera genes were negatively selected relative to HGP genes (dark line)
- Selection is less bad when focusing on genes from 2001 (most promising genes already sequenced by that point) (dashed line)
- No evidence of selection between Celera genes resequenced by HGP in 2002 relative to 2003 (dotted line)

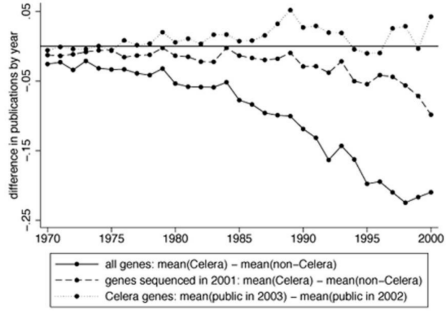


FIG. 3.—Investigating selection into Celera's IP. This figure provides three sets of descriptive statistics investigating the selection of genes into Celera's IP. The solid line (all genes) plots the difference in mean publications on Celera genes and mean publications on non-Celera genes in each year from 1970 to 2000. The dashed line (genes sequenced in 2001) plots the difference in mean publications on Celera genes and mean publications on non-Celera genes that were sequenced in 2001 in each year from 1970 to 2000. The dotted line (Celera genes) plots the difference in mean publications on Celera genes resequenced in 2003 and mean publications on Celera genes resequenced in 2002 in each year from 1970 to 2000. See the text and online Appendix A for more detailed data and variable descriptions.

Empirical strategies

These selection results motivate two empirical strategies

1. A panel strategy. Celera lost its IP when the HGP re-sequenced their protected genes. Thus the author can estimate for gene-year gt:

$$y_{gt} = \alpha + \delta_g + \gamma_t + \beta \cdot \text{Celeta}_{gt} + \varepsilon_{gt}$$

   Celera dummy transitions from 1 to 0 when Celera loses its IP. Gene-level fixed effects address selection concerns

2. Focusing only on Celera genes that were re-sequenced by HGP and taking advantage of variation in the timing of resequencing (2002 vs. 2003)

Panel results

- Column (3) has the gene fixed effects
- 45% fewer publications
- 22% fewer linked phenotypes

TABLE 4
PANEL ESTIMATES: IMPACT OF CELERA'S IP ON INNOVATION OUTCOMES

	(1)	(2)	(3)
A. Publications (mean = .244)			
Celera	−.160	−.121	−.109
	(.017)***	(.011)***	(.011)***
B. 1(known, uncertain phenotype) (mean = .381)			
Celera	−.163	−.160	−.083
	(.009)***	(.008)***	(.008)***
Year fixed effects	Yes	Yes	Yes
Indicator variables for year of disclosure	Yes	Yes	No
Number of publications in each year 1970–2000	No	Yes	No
Gene fixed effects	No	No	Yes
Observations	250,938	250,938	250,938

NOTE.—Gene-year-level observations. All estimates are from OLS models. The sample includes all gene-years from 2001 to 2009 (27,882 genes for 9 years implies $N = 250{,}938$ total gene-year observations). Robust standard errors, clustered at the gene level, are shown in parentheses. Celera: 0/1, =1 for a Celera gene. Indicator variables for year of disclosure: 0/1 indicator variables for the year the sequence for the gene was disclosed. Number of publications in each year 1970–2000: count variables for the number of publications on each gene in each year from 1970 to 2000. See the text and online App. A for more detailed data and variable descriptions.

* $p < .10$.
** $p < .05$.
*** $p < .01$.

Timing results

- Follow-on innovation lags for Celera genes made public in 2003 relative to those made public in 2002

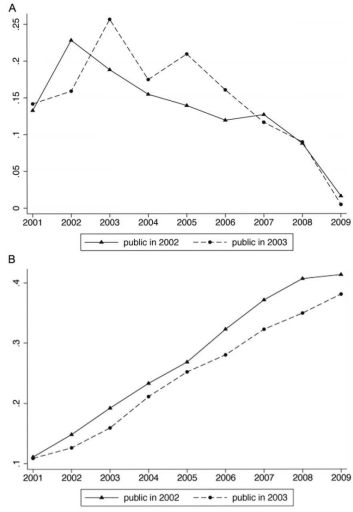


FIG. 5.—Average innovation outcomes for Celera genes by year, by year of resequencing by the public effort: *A*, Outcome variable: publications; *B*, outcome variable: 1(known/uncertain phenotype). These figures plot the descriptive statistics described in Section III.C. The sample includes all Celera genes. Means are shown separately for Celera genes that were resequenced by the public effort in 2002 ($N = 1{,}047$) and for Celera genes that were resequenced by the public effort in 2003 ($N = 635$). See the text and online online Appendix A for more detailed data and variable descriptions.

Does stronger IP lead to more innovation?

- Moscona (2022)
- Budish, Roin, and Williams (2015)

Does IP inhibit follow-on innovation?

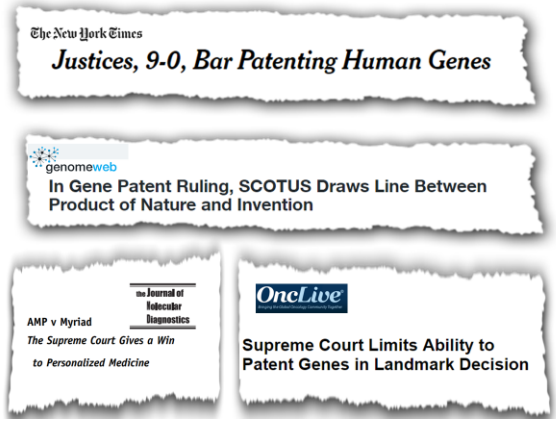
- Williams (2013)

Does disclosure matter?

- Sampat and Williams (2019)

What about gene patents?

- Prior to 2013 it was legal to patent genes
- Recall that Celera did not actually patent their genes. In fact, their patent applications were denied so they used a bespoke kind of IP...this will be important
- Sampat and Williams use a judge design to estimate the causal effect of patent protection on follow-on innovation
- I won't have time to talk about the whole paper today! But I want to highlight the conclusion...



No effect on follow-on innovation

- The authors find NO effect of patents on follow-on innovation
- Does this contradict the findings of Williams (2013)?
- Authors argue no: Celera's non-patent IP had worse disclosure. The gene sequence is easily visible in a patent! But with Celera, you had to pay for a CD-rom to access the data
- The disclosure function of the patent system is very important in this case
- If you want to restrict patenting, you should think about what firms will do instead