Revisiting the Entrepreneurial Commercialization of Academic Science: Evidence from "Twin" Discoveries

Matt Marx & David H. Hsu*

May 2020

Abstract: What factors shape the commercialization of academic scientific discoveries via startup formation? Prior literature has identified several contributing factors but does not address the fundamental problem that the commercial potential of a nascent discovery is generally unobserved and potentially confounds inference. We construct a sample of over 20,000 "twin" scientific articles, which allows us to hold constant differences in the nature of the advance and more precisely examine characteristics that predict startup commercialization. In this framework, several commonly-accepted factors appear not to influence commercialization. However, we find that teams of academic scientists whose former collaborators include "star" serial entrepreneurs are much more likely to commercialize their own discoveries via startups, as are more interdisciplinary teams of scientists.

Keywords: university technology transfer; entrepreneurship; technology commercialization; "twin" scientific discoveries.

---

* Marx: Boston University & NBER, mattmarx@bu.edu. Hsu: University of Pennsylvania, dhsu@wharton.upenn.edu. We are grateful for feedback from Christine Beckman, Kristoph Kleiner, and participants of the Wharton Technology & Innovation Conference; the Global Entrepreneurship & Innovation Conference; the *Strategy Science* Conference, West Coast Research Symposium at Stanford; the Crete Workshop on Innovation & Creativity; as well as department seminars at Washington University in St. Louis, Cornell University, and KAIST. We thank Chris Ackerman, Rafael Castro, Andrea Contigiani, and Luming Yang for excellent research assistance. We also thank Guan-Cheng Li for data on patent-to-paper citations, Michael Ewens for his USPTO/VentureSource concordance, and Kyle Myers for his USPTO/Crunchbase/SBIR concordance. We acknowledge research support from Boston University and the Mack Center for Innovation Management at the University of Pennsylvania. Errors and omissions are ours.

1. Introduction & Motivation

What factors shape the commercialization of academic scientific discoveries via startup formation? The technologies underlying many successful companies—including Google's PageRank search algorithm, E-Ink's "electronic paper", RSA's cryptography algorithm, and Genentech's recombinant growth hormone—were discovered by scientists at universities who then commercialized them via startups. Consider Amnon Shashua, professor of computer science at Hebrew University, who published dozens of articles applying computer vision to traffic safety, including "Forward Collision Warning with a Single Camera." Shashua might have left his work in the public domain for others to notice and possibly exploit but instead chose to self-commercialize this research by co-founding Mobileye, which supplies the driver-assistance systems in many vehicles and became Israel's largest startup acquisition when sold to Intel for $15.3B.

With universities increasingly concerned with economic development alongside their longstanding teaching and research missions, scholars have sought to better understand academic entrepreneurship. Rothaermel, Agung & Jiang (2007) and Markman, Siegel & Wright (2008) catalog 175+ such papers. Commercialization can take several forms including technology licensing to established firms—and we do not claim that a startup is always the optimal vehicle—but the literature has highlighted several reasons to understand new venture formation from academia. First, technologies developed in university labs are typically more embryonic than their industrial lab counterparts (Jensen & Thursby, 2001). As a result, absent new venture development, these discoveries may go uncommercialized (Hsu & Bernstein, 1997). Second, commercializing discoveries via venture formation addresses the changing nature of careers in academic science. STEM doctoral degree awardees in the U.S. have long exceeded the number of available academic jobs (Cyranoski, et al. 2011) while there has been a steady increase in university spinoffs. Because graduate students are instrumental in academic entrepreneurship (Hayter, Lubynsky & Maroulis, 2017), there are scientific labor market implications of commercializing science. Finally, from an economic development standpoint, startups are disproportionately involved in job growth (Haltiwanger, Jarmin & Miranda, 2013). Academic ventures tend to locate near prominent research scientists (Zucker, Darby & Brewer, 1998), so regional growth may be spurred by venture formation.

Work on the commercialization of academic research via startup formation has focused on two sets of antecedents. A first view, which we call the *resource munificence* perspective, claims that entrepreneurial opportunities proceed to commercialization based on resources available, often within a given geography. Resources could include financial capital (Samila & Sorenson, 2011) and know-how, spanning technical as well as commercial domains, and may take place at various levels of analysis including groups (e.g., workshops) and even institutions (e.g., university collaborative relations with private enterprises).

Financial capital for developing entrepreneurial opportunities such as from venture capitalists is thought to be particularly sensitive to geographic co-location. Resources flow based on researcher or institutional prestige, so this literature has also examined the role of status (Stuart, Hoang & Hybels, 1999).

A second perspective, which we term the *discovery team composition* view, highlights the configuration and social context of the team producing the scientific advance. This branch of literature suggests that scientific teams with exposure to peers with experience in commercializing science can have substantial effects on the propensity of engaging in entrepreneurship due to awareness, demonstration effects, professional legitimization, and experience with commercialization (e.g., Stuart & Ding, 2006). Team composition itself can also impact entrepreneurial opportunity recognition and commercialization outcomes such as through social networks and experience (e.g., Baron, 2006).

The literature review by Rothaermel, Agung & Jiang (2007) summarizes these two categories (see their Figure 7 on p. 761). These theories have only rarely been jointly assessed, and even when examined together suffer from a fundamental empirical problem impairing the entire literature on the antecedents of academic entrepreneurship: unmeasured *latent commercializability*. By this we mean that each scientific discovery has a distinct level of commercial potential, which may be difficult to discern (and is perhaps unclear even to the participants). Indeed, the literature on academic commercialization frequently characterizes academy-originated technologies as "embryonic" (e.g., Jensen & Thursby, 2001), which compounds the difficulty of ascertaining eventual suitability to the commercial market.

Researchers have rarely attempted to control for latent commercializability. Azoulay, Ding & Stuart (2007) construct such a measure for the life sciences based on keywords assigned by PubMed that overlap with words in patent applications. But even within a set of keywords, there may be vast differences in commercial potential. We instead tackle this confound by building on the Bikard & Marx (2019) method of analyzing "twin" scientific discoveries, though we dramatically scale up their effort to include all fields of science over a 60 year period, studying the antecedents of startup commercialization among more than 20,000 twin discoveries. This approach allows us to examine the resource munificence and discovery team composition views on a comparative basis while taking off the table technology differences (and their latent commercializabilty). We therefore mitigate two inference problems plaguing the prior literature. First is the issue of spurious correlations resulting from unaccounted-for differences in commercial potential. Second, even if estimated correlations are not spurious, without controlling for commercial potential it is difficult to discern the degree to which results are due to selection.[1]

[1] Consider the finding of Stuart & Ding (2006) that prior association with a professor who has entrepreneurial

Our study illustrates how *not* accounting for latent commercializability can dramatically alter inferences regarding the antecedents of academic entrepreneurship. We begin with a population-level analysis of published academic findings (more than 38 million academic articles in the Web of Science (WoS), 1955-2017) *without* controlling for technology commercial potential. Through this analysis, we replicate prior results on the importance of resource-munificence and discovery-team-composition for entrepreneurial commercialization. When we then account for commercial potential by analyzing twin discoveries, however, although we confirm the discovery-team composition view, we do not find evidence for the resource-munificence perspective. Our empirical approach ensures that these differences are not due to different variable definitions or data sources.

Moreover, controlling for latent commercialization refines our understanding of the role of discovery-team composition. Only when examining "twin" discoveries can we conclude that prior results regarding peer effects in academic entrepreneurship are not driven solely by (for example) mentors directing the selection of projects with commercial promise. Rather, *even when considering the same scientific discovery,* scientists with entrepreneurial peers are more likely to commercialize. Similarly, although in the cross-section we confirm past results suggesting that interdisciplinary research teams might commercialize less due to coordination costs, this result reverses when accounting for latent commercializability. When more interdisciplinary teams of scientists develop the same scientific discovery as less diverse teams, they are more likely to commercialize via startup formation.

We proceed by describing the construction of the twins dataset in Section 2. In section 3 we compare results in the cross-section with the twin-discovery method. In Section 4, we discuss how our results revisit the existing literature. A concluding Section 5 reviews our contributions and highlights limitations.

2. Empirical approach

As noted above, unmeasured technological differences in academic discoveries and their associated commercial potential is a critical confound in the literature. The ideal experiment to assess when academic researchers commercialize their discoveries via startups would involve random matching of researchers and discoveries, which seems impractical in a university environment as few scholars would consent to being assigned projects and/or colleagues. Instead, we take advantage of the fact that different research teams sometimes make identical (or very similar) discoveries, which we label "twins," thereby

---

experience positively predicts the focal professor also doing so. This finding might be driven by learning about what it means to be an entrepreneur, as typically put forth in work on peer effects and entrepreneurship (Nanda & Sorensen, 2010). In the case of academic entrepreneurship, however, it could be due to the influence of a mentor on project selection. That is, researchers intentionally select scientific avenues of inquiry with higher commercial potential. Without being able to account for latent commercializability, it is difficult to tease this mechanism apart.

allowing us to hold constant technological differences in shaping startup commercialization. We build on an existing methodology of identifying scientific twins. This design allows us to estimate counterfactuals such as: if a given scientific advance had been made in two different entrepreneurial ecosystems, is there a tendency for the advance made in the more munificent financial environment more likely to proceed to entrepreneurial commercialization? In this section, we outline the method for assembling the "twins" dataset and contrast results with those from the population.

We adopt and expand upon a method of identifying twins based on common citation patterns (Bikard & Marx, 2019). Citations act as a window into the allocation of credit within the scientific community (Cozzens, 1989), so one can infer co-discovery status from papers with distinct authorship but similar citation patterns. Although co-discoveries are uncommon in the social sciences, they are more common in the "hard" sciences as many research teams are chasing the same scientific frontier. Journal editors may appreciate the opportunity to publish concurrent scientific advances—indeed, twins often appear back-to-back in the same issue of the same journal—as a reaffirmation of the accuracy of the finding. Indeed, Bikard & Marx (2019) verified the method by hiring 10 postdoctoral researchers to manually review dozens of twins that had been identified automatically, with no false positives reported.

We begin by replicating exactly the methodology of Bikard & Marx (2019), finding all pairs of papers that satisfy five conditions: 1) published no more than a year apart; 2) zero overlap between the authors; 3) are cited at least five times; 4) share 50% of forward citations; 5) jointly cited by at least one other paper (i.e. in the same reference list). Our methodology departs from theirs, however; instead of limiting our analysis to articles from the top 15 scientific journals between 2000 and 2010, we apply these criteria to the entire WoS from 1955-2017. Doing so yields a set of *potential* twin discoveries embodied in 40,392 papers. The next step in the methodology is to determine which of the potential twin discoveries are cited not just jointly (i.e., in the same reference list) but *adjacently* (i.e., within the same parenthesis). Adjacent citations suggest that forward-citing researchers are unable to attribute the discovery to a single paper, with the references listed within the citation parenthesis receiving co-attribution.

Identifying adjacent citations involves inspecting the text of papers that jointly cite what may be twin discoveries. For the 40,392 *potential* twin papers, both appear in the reference lists of more than 1.2M papers. Retrieving all such papers is impractical, as many if not most published articles reside behind paywalls and are inaccessible at scale. However, PDFs of many papers are freely available—sometimes in draft form—and have been indexed by Google Scholar (GS). Although GS does not support bulk downloads, over a period of 19 months we were able to retrieve approximately 280,000 publicly-available, non-paywalled PDFs of to the 1.2M papers that jointly cited our potential twin discoveries. For 29,257 of the 40,392 potential twin discoveries, we were able to determine whether they were adjacently cited by the PDFs that cited both of them. Of those, we found that 23,851 *potential* twin papers were indeed cited adjacently.[2] These comprise our population of twin discoveries, which should have similar latent commercial potential among twins.[3] Appendix A provides more detail on the twin discoveries, which hail from more than 3,000 academic institutions in 106 countries and span more than 200 scientific fields.

2.1 Dependent variable: entrepreneurial commercialization of scientific discoveries

Our dependent variable indicates whether academic researchers commercialize their discoveries via a startup. To our knowledge, a large sample of academic scientific discoveries commercialized via startups has not been previously assembled. Several studies of technology transfer have tracked out-licensing or other forms of commercialization more generally, not necessarily via new venture formation (see Rothaermel, Agung & Jiang (2007) for a review). There have also been academic institution-specific studies of new venture formation (e.g., Kenney & Goe, 2004; O'Shea et al. 2005) as well as sector-specific studies of commercial science, most notably in the biotechnology industry (e.g., Zucker, Darby, & Brewer, 1998; Stuart & Ding, 2006). Our aim, however, is to identify entrepreneurial commercialization of scientific discoveries at scale, spanning academic institutions, industrial sectors, and geography. Aside from the benefit of algorithmically assembling a large empirical sample, our method allows us to directly trace new ventures all the way back to a particular scientific advance, a feature novel to the literature.

We measure entrepreneurial commercialization in two ways. First, we detect entrepreneurial commercialization via patent-paper pairs ("PPPs") (Murray, 2002) where the patent is assigned to an entrepreneurial venture. The premise is that while scientific publications are the typical currency of academia, patents and their associated legal protection are valued much more in the commercial domain, and specifically by venture capitalists (Hsu & Ziedonis, 2013). Our algorithmic effort is therefore aimed at identifying patents which are granted to entrepreneurial ventures which cover the same or similar scientific advance in which there is overlap in authors. We start by finding the subset of academic discoveries that are cited by patents (Marx & Fuegi, 2020) and check for overlap between the authors of the paper and the inventors named on the patent. Article authors and patent inventors are compared individually, with an overall match score computed according to a) whether the surname is an exact

---

[2] Of the 23,851 twins identified, multiple adjacent citations were found for 62%. This count of adjacent citations is a lower bound, as we could retrieve only 280,000 of the 1.2M papers where both twins are in the reference list. If it had been possible to inspect all 1.2M papers, we likely would have found multiple adjacent citations for more twins. In Table 8, we drop the twins established via a single adjacent citation, yielding similar results.

[3] Twin discoveries are not randomly distributed. In unreported statistics, we find that twin discoveries tend to be somewhat more recent than papers on average and are published in higher impact factor journals. They are, however, cited at similar rates to the full WoS corpus. Moreover, the scientific teams on twin papers are no more or less interdisciplinary than scientific teams on average. Twin papers are much more likely to have a "star" commercializer among the authors (or the authors' past collaborators), however.

versus fuzzy match; b) frequency of the surname in the WoS and the patent corpus; and c) whether the middle initial matches (more details are provided in Appendix B). A weighted average of author/inventor overlap is computed to yield an overall article/patent match score.[4] However, not every patent-paper pair represents entrepreneurial commercialization. For example, one or more scientists on a paper may cooperate with an established firm to commercialize the discovery. We thus subset the list of patent-paper-pairs to those assigned to startups, as determined from VentureSource and Crunchbase.

Our second method involves U.S. Small Business Innovation Research (SBIR) grants. The SBIR program is targeted at encouraging "domestic small businesses to engage in federal research and research & development that has the potential for commercialization" and has awarded non-dilutive funding in excess of $45B since the program was initiated in 1982 (www.sbir.gov/about). We interpret pursuing SBIR funds as an indicator of commercialization aspirations. Note that the SBIR channel of identifying commercialization attempts does not rely on observed patenting: this may be an important complement to the PPP measure, as Fini, Lacetera & Shane (2010) suggest that only about a third of businesses started by academics are based on patented inventions. Moreover, the literature's reliance on patent data is likely related to the fact that our understanding of technology commercialization heavily relies on the biotechnology industry (see, for example, the discussion in Hsu, 2008), as patents are well-understood to be important as an appropriation method in that industry (e.g., Levin, et al. 1987). As with patent-paper pairs, we calculate the pairwise overlap between scientists on a focal article and either the primary contact or principal investigator of SBIR awards up until five years after the publication of the article. If multiple SBIR awards have identical author-overlap scores, we break ties with temporal proximity.

Overall, we find 139 academic articles that were commercialized via PPPs assigned to startups and 89 that were commercialized via SBIR awards, for a total of 228 entrepreneurial commercialization events. Appendix B also provides validation of the measure, confirming via web research a stratified random sample that both PPPs and overlapping SBIR grants truly reflect instances of a startup commercializing an academic discovery with the involvement of one of the original scientists. In short, we verified 20 out of 20 of the PPP-based commercialization events, and 19 out of 20 SBIR-based events.

2.2 Explanatory variables

Our explanatory variables fall into the aforementioned categories of resource munificence and

[4] If the authors of an article have an identical overlap score with the inventors on multiple patents, ties are broken in two steps. First, the PPP closest in time is retained. Second, if two patents in the same year form pairs with the same paper, we further resolve ambiguity via cosine similarity between the abstract of the article and the summary text of the patent, computed using Term Frequency * Inverse Document Frequency.

discovery-team composition. Resource munificence is often tied to geography (Samila & Sorenson, 2011), so we constructed a lagged count of venture capital investments in the same postal code as the discovery. Resources also often accrue to high-status actors, so our second and third measures of munificence reflect the prestige of the discovery team and their associated institutions (Stuart, Hoang & Hybels, 1999). Each of these variables is calculated as a count of publications (per author, or per institution) in the same scientific field as the focal paper. WoS assigns each article to one of 251 scientific fields.[5]

Regarding discovery team composition, a first variable measures the interdisciplinarity of the scientists. This is calculated as one minus the Herfindahl-Hirsch index of scientific fields for articles written by the authors. If all scientists on the focal article published all of their papers in the same scientific field, this variable is zero. A second explanatory variable measures whether the *previous* collaborators of authors on the paper include a 'star commercializer.' This variable is reminiscent of Stuart & Ding's (2006) measure of the number of prior collaborators who served as founders or advisory-board members of startups that filed for an IPO, but our measure differs in three ways. First, we measure involvement with early-stage ventures, not just those that complete an IPO. Second, instead of summing all instances of entrepreneurial involvement, we focus on "star" commercializers (above the 75^{th} percentile of entrepreneurially-commercializing academic scientists in the year of the scientist's most recent collaboration (similar results are obtained at the 50^{th} or 90^{th} percentile). Third, instead of focusing on individual scientists, we check whether *any* scientist on the discovery team had previously collaborated with a star. Additional characteristics of star commercializers are available in Appendix C. As a third team-composition covariate, we control for whether any of the authors on the paper is herself a star commercializer.

2.2 Controls

Given that our dependent variable depends on a name-matching algorithm, articles with more authors might mechanically have higher overlap scores. We therefore control for the number of scientists on each paper. Similarly, given that our PPP measure of commercialization relies on patent-to-paper citations, we control for the number of citations to an article from (industry) patents, regardless of author overlap.

2.3 Empirical specification

Following epidemiological twin studies (Carlin et al., 2005), we estimate the likelihood of

---

[5] For institutions in North America, we also have technology-transfer related variables from the Association of University Technology Managers and compute models limited to institutions where such variables are available. However, because the AUTM data rely on respondent survey responses which are self-reported and because of the limited (domestic) coverage of the data only among some association members, we do not report these models.

commercialization using fixed effects for papers reporting a twin discovery. The regression equation is:

$$ENTCOMM_{ij} = \alpha_0 + \alpha_1 RESOURCE\_MUNIFICENCE_i + \alpha_2 DISCOVERY\_TEAM\_COMPOSITION_i + \alpha_3 X_i + \gamma_j + \varepsilon_{ij}$$

where *j* represents the twin discovery and *i* represents a paper reporting the twin discovery. $ENTCOMM_{ij}$ captures whether the focal article was commercialized by a startup. $RESOURCE\_MUNIFICENCE_i$ captures variables related to local venture capital investments and institution/discovery team prestige measures. $DISCOVERY\_TEAM\_COMPOSITION_i$ variables measure the interdisciplinarity of the scientific team and whether the scientists on a given article had previously collaborated with a "star" entrepreneurial commercializer. X_i represents controls for the number of authors and count of citations from industry patents. Finally, γ_j is a fixed effect for the twin discovery. We estimate this equation using linear probability models (LPM) and robust standard errors. Following Beck (2015), in the robustness checks of Table 8 we also estimate conditional logit models, which exclude any twin discovery where neither (or both) of the twin discoveries is commercialized.

3. Results

Before proceeding to analyze the entrepreneurial commercialization of academic science via our twin-discovery methodology, we first present cross-sectional estimates of the correlation between our covariates and our dependent variable using *all* known scientific publications. Our motivation for doing this is twofold. First, as noted above despite hundreds of papers on related topics only rarely have resource-munificence and discovery-team composition factors been jointly considered. Second and more importantly, to the extent that our findings using the twins methodology differ from prior work, one might suspect that this is due to differences either in variable construction or in the nature of the data analyzed. As noted above, past work has frequently analyzed patented inventions (or invention disclosures), or a set of projects from a small number of universities. By comparison, we analyze all academic articles. In what follows, we replicate prior results using the universe of scientific publications. After addressing technology and latent commercializability differences via the twins method, the magnitude and sometimes the sign of results change.

3.1 Population-level results without controlling for latent commercializability

We proceed by considering all academic articles in the Web of Science, 1955-2017. The resulting sample of 38,715,150 articles contains 9,654 instances of entrepreneurial commercialization (0.02%). Of these, 8,361 were found via patent-paper pairs and 2,996 via SBIR grants. (17 were in both categories.) Descriptive statistics are found in Table 1. In Table 2, we estimate the above regression equation, replacing twin-discovery fixed effects with year, country, and scientific field. Column (1) of Table 2 includes control variables only, demonstrating the importance of controlling for the number of authors as well as the count of citations from industry patents. Following the literature, columns (2-7) examine each explanatory variable independently and finds results consistent with prior work. All covariates are considered jointly in column (8), with weaker statistical significance for institutional prestige.

[Table 1 and Table 2 about here]

Given the sample size, it is unsurprising that most correlations are statistically significant and so we focus on estimated effect magnitudes. These are calculated by estimating the predicted probability of commercialization given a one standard-deviation increase in the explanatory variable, holding all other covariates at their means. We first consider resource-munificence factors. Using estimated coefficients from column (8), an increase of 5 citations to the authors (a one standard deviation increase) of a paper is associated with an 8.3% rise in entrepreneurial commercialization. Two additional venture capital investments in a given zip code is associated with a 15.2% increase in commercialization. These results suggest, consistent with prior research, that prestige and the availability of financial capital facilitate commercialization, thus supporting the resource munificence view.

Moving to discovery-team composition factors, we likewise confirm prior results in the cross-section. First, we confirm Stuart & Ding's (2006) findings regarding the strong role of both having a star commercializer either on the discovery team (70% increase) or among one of the team's prior collaborators (83% increase). Second, the 2% decrease in commercialization for a one-standard-deviation (0.29) increase in interdisciplinarity is reminiscent of Bercovitz & Feldman's (2008) negative association between researchers from different departments a) patenting their discoveries b) licensing those patents c) obtaining revenue from licensed patents. By replicating prior results from the literature based on the full population of academic articles since the 1950s, we hope to allay concerns that differences we find via our twin-discovery method are not due to data or variable construction differences.

3.2 Main results using twins methodology

Table 3 contains descriptive statistics and correlations for our twin-discovery methodology. Table 4 compares twin discoveries that were entrepreneurially commercialized versus not on a descriptive basis. Discoveries commercialized by startups have more scientists, more prestigious scientists, and more interdisciplinary scientific teams. Commercialized discoveries have about 40% more citations from industry patents and, interestingly, are from institutions with somewhat lesser prestige. Perhaps the most dramatic univariate difference is in the share of discoveries that have a star commercializer among the scientists themselves or their prior collaborators. Half a percent of non-commercialized discoveries have a star on their team, compared with nearly 10% of commercialized discoveries. We see a similar pattern for prior ties to star commercializers. As in the population-level analysis, commercialized discoveries are also located in postal codes with more venture capital investments.

[Table 3 and Table 4 about here]

We begin our regression analysis in Table 5. In column (1), twin papers from larger teams of scientists are more likely to be commercialized via startups. This result underscores the importance of controlling for the number of authors, as our measure relies on name overlap between article authors and either patent inventors or SBIR recipients. Column (1) also fails to precisely estimate the relationship between citations from patents assigned to established firms, which is reassuring as one might be concerned that our PPP dependent variable could be conflated with articles simply being cited more often by patents.

In Columns (2-4) we assess the role of resource munificence using the twins analytic strategy. In column (2), when accounting for latent commercializability the prestige of the authors does not appear to materially impact entrepreneurial commercialization. Although the estimated coefficient is positive, it is imprecisely estimated with a small t-statistic. The prestige of the institution, analyzed in column (3), also does not appear to play a crucial role in commercialization. Finally, we find that the entrepreneurial commercialization of academic science is not explained by the number of local venture-capital investments (column 4) when accounting for latent commercializability. While we discuss in more detail how our results relate to findings from the literature in Section 4, one explanation for the difference in the local VC effect may be the movement of researchers with higher commercial potential projects (or an applied orientation more generally) to particular geographic regions.

[Table 5 about here]

In columns (5-7) we apply the twins methodology to assess whether controlling for latent commercializability can also rule out findings related to the composition of the scientific team responsible for the discovery. When interpreting these measures, we caution that our variables do not necessarily reflect the composition of the founding team of the startup but rather the "discovery team" that pioneered the original scientific finding in academia. Under our definition, one or more of the authors was indeed involved with the startup (either as a patent inventor or an SBIR recipient). Column (5) indicates that discoveries where the scientific team is more interdisciplinary are more likely to be commercialized by startups than when the discovery team is homogeneous with respect to scientific field. Column (6) shows that discoveries are more likely to be commercialized via startups when one of the authors is a star commercializer. In column (7), we find similar confirmation for discovery teams where one of the authors has previously worked with a star commercializer. All of the foregoing covariates are included in column

(8), which strengthens statistical significance on the estimated coefficient for interdisciplinarity.

Using estimated coefficients from column (8), a one-standard-deviation increase in interdisciplinarity (0.27) corresponds to a 2.7% increase in the likelihood of commercialization by a startup. The presence of a star commercializer among the scientists' past collaborators is associated with a 4.1 percentage-point increase, and having a star commercializer among the authors themselves predicts a 9.5 percentage-point rise in the likelihood of commercialization. These findings are plotted in Appendix D.

3.3 A deeper dive into discovery team composition effects

In Table 6, we explore which forms of interdisciplinarity matter for entrepreneurial commercialization. In column (2), we replace the interdisciplinarity variable from Table 5 (shown in column (1) for comparison) with a count of unique primary scientific fields among the authors on the paper. By "primary" we mean the scientific discipline in which each author publishes most often. The positive, statistically-significant estimate of the associated coefficient suggests that having scientists from a variety of scientific fields is important, not just having a set of scientists from the same discipline who also work relatively often in other areas. The magnitude of the estimated coefficient on the simple count of scientific fields represented is considerably smaller than the more subtle measure of interdisciplinarity among all authors, however. This suggests that merely having more scientific fields represented does not fully explain the findings in column (1).

[Table 6 about here]

Column (3) suggests that the optimal discovery team for commercialization requires more than a team of specialists. The covariate in this column counts the number of scientists who fully specialize (i.e., all of their work is published in a single scientific field). If specialization alone were critical to the commercialization process, we might expect this coefficient to be significant, but it is does not appear so. Or, perhaps it is the case that an ideal configuration would combine a set of specialists with a boundary spanner. In column (4), we calculate each scientist's individual level of interdisciplinarity and then enter as a covariate the difference between the most interdisciplinary scientist and the mean of the team. The negative coefficient suggests that having one scientist much more interdisciplinary than most does not facilitate commercialization. Taken together, columns (2-4) suggest that the ideal composition is a well-rounded team of well-rounded scientists. As we discuss further in section 4.3, this result is consistent with Baron's (2006) claim that "opportunity recognition can be enhanced by providing potential entrepreneurs with a very broad range of experience…the broader this experience…the more likely the entrepreneurs will be to perceive connections between seemingly unrelated events or trends."

The remaining column of Table 6 verifies that it is the presence of a star *commercializer* among the scientists' past collaborators that explains the patterns in Table 5 and not simply an association with a prominent researcher. In column (5), we replace the star commercializer variable with an indicator for a star *scientist*—those whose citation count per article (in a five-year window following publication) was in the 99th percentile—among one's past collaborators. Following Zucker et al. (1998), we would expect this coefficient to be statistically significant, but it is imprecisely estimated. We conclude that star commercializers rather than star scientists facilitate the entrepreneurial commercialization of science.

3.4 Heterogeneity

In Table 7 we attempt to understand whether the results from Table 5 are driven by geography, scientific field, or time periods, with the caveat that splitting the sample along these dimensions may yield noisier estimation due to statistical power. Column (1) repeats column (8) of Table 5 for convenience. Columns (2-4) split the sample of twins into three groups: 1) both twins are in the U.S., 2) neither of the twins in the scientific discovery is in the U.S., and 3) the scientific discovery contains a twin from the U.S. as well as a twin from outside the U.S. The majority of twins are "mixed"—i.e., in the third category. U.S.-based twins are more often commercialized when the discovery teams are interdisciplinary, as well as when those teams include a star commercializer, as is visible in column (2). Prior collaboration with a star commercializer appears to play a role only for twins containing non-U.S. articles (columns 3 and 4).

[Table 7 about here]

Regarding scientific field, much of our collective empirical knowledge on commercialization draws on the field of biotechnology, perhaps due to data availability reasons, as previously discussed.[6] In column (5), having a star commercializer on the paper predicts commercialization of biotechnology discoveries, but neither interdisciplinarity nor prior association with a star does. Of course, the life sciences are not limited to biotechnology; in column (6) we analyze non-biotechnology papers in life sciences[7] and find that interdisciplinarity plays a key role. Prior association with a star commercializer plays a role as well. Finally, in column (7) we analyze non-life-sciences papers, of which there are more than either of the other categories. We find no statistically significant covariates.

---

[6] We define biotechnology as the following categories from the Web of Science: Biochemistry & Molecular Biology, Cell Biology, Microbiology, Cell & Tissue Engineering, Biology, and Biology & Applied Microbiology.

[7] Life sciences not including biotechnology includes OECD categories 1.06 (Biological Sciences), 3.01 (Basic medical research), 3.03 (Health Sciences), and all papers in Multidisciplinarity Sciences (of which 99% are in Nature, Science, or PNAS). The OECD categories are mapped to the WoS categories using a Clarivate crosswalk http://help.incites.clarivate.com/inCites2Live/5305-TRS/version/default/part/AttachmentData/data/OECD%20Category%20Mapping.xlsx.

Columns (8) and (9) attempt to situate these results temporally, splitting the sample into papers published before and after the year 2000. It appears that interdisciplinarity played a significant role only for papers published in the 20th century. Moreover, we observe a shift from reliance on having a star among the authors of the paper earlier (column 8) to prior association with a star (column 9).

3.5 Robustness and placebo tests

Table 8 contains robustness checks and placebo tests, column (1) again repeating column (8) of Table 5 for comparison. Column (2) re-estimates column (1) in a conditional logit framework (see Beck, 2015). Because the maximum likelihood estimator drops groups without variation in the dependent variable, the fixed effects for each twin discovery reduces the number of observations. Statistical significance is reduced somewhat for the interdisciplinary result ($p<0.07$). In column (3), we compare the logistic and OLS specifications by manually limiting observations to the set of twin discoveries with variation in the outcome variable. When doing so, OLS results closely resemble that of the logit estimates.

As noted above, we were able to check for adjacent citation in only 280,000 of the 1.2 million papers that jointly cited potential twins. Although for 38% of our twins we found only a single adjacent citation, that figure may be understated as we could not inspect three-fourths of the co-citing PDFs. We thus check that our results are robust among twins that we can confirm were adjacently cited multiple times. Column (4) confirms the results are similar, with a slightly smaller t-statistic for the interdisciplinarity coefficient.

Finally, in column (5) of Table 8 we perform a placebo test by randomly generating values for the dependent variable. Doing so yields no statistical significance on any covariates. In unreported results, this placebo test also fails if the distribution of the randomly-generated dependent variable matches that of the actual dependent variable (i.e., less than 1% of papers are commercialized by startups).

[Table 8 about here]

4. Calibration with the academic entrepreneurship literature and future study directions

We now discuss how the population-level, cross-sectional estimates contrast with those from the twin study, and how these results compare with those reported in the literature.

4.1 Entrepreneurial ecosystem resource munificence

The munificence of resources required to commercialize entrepreneurial discoveries has been a frequent subject of inquiry. For example, Zucker, Darby, & Brewer (1998) report that U.S. states with more academic scientists who have outsized academic output ("stars") are home to more biotechnology startups, particularly in the nascent phase of industry development. Their study does not establish direct linkages between startups and academic scientists, but their findings have generally been interpreted to suggest that the localized knowledge of highly productive scientists may be important in the geography of entrepreneurial commercialization. Such an interpretation is consistent with what we see in the population-level cross section of all papers from the Web of Science in Table 2, even though the Zucker et al. sample and variables differ substantially from ours. However, when accounting for latent commercialization via our twins methodology, we no longer find support for the role of prominent scientists. It seems that academic prominence alone does not influence the commercialization process.

Similarly, an extensive literature has found a connection between the local munificence of venture capital and the founding of new firms (Samila & Sorenson, 2011). We see a similar association in the cross-section. It seems unsurprising that discoveries distant from sources of venture funding would not be commercialized, but is the mechanism based on a treatment or selection effect? Samila & Sorenson utilize an instrumental-variables approach to adjust their estimates to account for the possible self-selection of venture capital to geography, so it seems that venture capital has a causal effect on entrepreneurship in general. The fact that we fail to find any connection once applying the twins methodology raises the question of whether this effect extends to academic entrepreneurship. It could instead be that discoveries in close proximity to sources of capital simply have more commercial potential, raising the possibility that commercially-minded academic scientists select into regions with locally-available financial capital.

Note that both the Zucker, Darby & Brewer (1998) and Samila & Sorenson (2011) papers are set in the US and (largely) in the time period before the year 2000. While it is difficult to directly compare our twins results on commercializing science with these papers (the outcome variables and of course the samples differ), we do not find evidence for the local knowledge capital or VC effect in our analysis of sample heterogeneity. This suggests that further research on the role of resource munificence while taking into more careful account the nature of technical opportunities produced in the local geography, especially as related to academic entrepreneurship, may be warranted.

4.2 Discovery team composition: "star" commercializers

Whereas we failed to find support for resource munificence when accounting for latent commercialization, we affirm the role of star commercializers (Stuart & Ding, 2006) both in the universe of scientific publications (cross-sectional analysis) and when using the twins methodologies. In addition to reconfirming prior findings, by controlling for latent commercialization we show that peer effects are not limited to project selection (i.e., students of star faculty commercializers pick research topics that are more commercializable). Rather, even when confronting the same scientific discovery, those with exposure to entrepreneurial peers are more likely to commercialize.

Our findings are consistent with broader work on peer effects, such as Nanda & Sorenson (2010), who use Danish data from 1980-1997 and find a positive immediate workplace peer effect on entrepreneurial entry. While it is difficult to directly compare our results to these studies due to different samples (among other things), the operative mechanisms behind the positive peer effects are demonstration effects ("if I can be an entrepreneur, you can as well") and information/resource effects ("in order to be credible to investors in this space, you have to have more than a product prototype"). Our heterogeneity results are in line with the literature that affiliation with commercialization stars may be a more general worldwide phenomenon and not idiosyncratic to one institutional environment or time period, nor driven as we previously concluded by detailed controls for technological differences and latent commercializability.

It is important to stress that our results should not be interpreted as *causal* evidence for peer effects in academic entrepreneurship. Although it is important to hold constant the nature of the discovery, our results do not address the endogeneity of discovery team formation. For example, perhaps the original scientist recruits a star commercializer to join the discovery team if s/he senses a startup opportunity. Although we know who the corresponding author is on each paper, we would need to know the order in which the team was assembled (not just order of authorship) in order to rule out this possible alternative. What we can rule out by controlling for latent commercializability is that project selection was influenced by prior exposure—i.e., pushing scientists toward projects that are more applied and thus have greater commercial promise. Moreover, Lerner & Malmendier (2013), using randomized graduate business student section assignments between 1997-2004, find peer effects dampen entrepreneurial entry (but that those who enter are more successful, suggesting that peers may provide information allowing focal individuals to "properly" assess their prospects). Their negative peer effect seems to stem from a "screening" function which may serve as a check to would-be entrepreneurs who may not have realized the many obstacles to successful entrepreneurial entry. While we do not study commercialization performance, our heterogeneity results in comparison to this study suggests that perhaps industry conditions outside of the biotechnology and life sciences may curtail the role of (star commercializer) peer effects. Consistent with the Lerner & Malmendier (2010) study, we do not find a general peer effect outside of the biotechnology and life science contexts. There may be less of a compelling peer screening function outside of the specialized expertise in the life and health sciences.

While the effects of peers on (academic) entrepreneurial starts is therefore not settled, especially as there are more than a few differences in the research contexts across the studies, we believe that a fruitful path forward is tying peer effects to the *specific* entrepreneurial opportunity. Note that in contrast to the prior peer effects literature which treats such opportunities as unspecified and unmeasured, our study begins the process of specifying the (scientific) advance giving rise to a potential entrepreneurial opportunity. At the same time, it will be important to pay particular attention to possible differences in the peer effect process in the life and health sciences as compared to other scientific sectors. Such an empirical strategy holds the promise of revealing more on how and why the nature of peer effect interaction matters for entrepreneurial commercialization.[8]

4.3 Discovery team composition: interdisciplinarity

Finally, we consider the interdisciplinary nature of discovery teams. Interdisciplinarity has been frequently studied (Leahy, Beckman, & Stanko, 2017), though rarely in the context of entrepreneurship. The most relevant articles of which we are aware are Berkovitz & Feldman (2008) and Kotha, George, & Srikanth (2013), both of which examine the licensing of university invention disclosures as opposed to startup formation. These authors find that more departments spanned by the discovery team (which they interpret as coordination costs) is negatively associated with a lower likelihood of licensing (though the effect reverses for a squared term indicating a curvilinear effect). Our cross-sectional results are largely similar, showing that more interdisciplinary teams are somewhat less likely to commercialize via startups.

When applying the twins methodology, however, the sign of the estimated coefficient reverses. That is, when a more interdisciplinary team develops a highly similar invention as a less interdisciplinary team, it is more likely to commercialize via startup formation. By controlling for latent commercialization, we take off the table the role of coordination costs and project selection. Having overcome these costs, there are at least two reasons to think that interdisciplinary teams are more likely to commercialize a given discovery. First, a more diverse set of perspectives among the original scientists may improve opportunity recognition due to varied inputs. Baron (2006:117) claims that "opportunity recognition can be enhanced by providing potential entrepreneurs with a very broad range of experience…the broader this experience…the more likely the entrepreneurs will be to perceive connections between seemingly unrelated events or trends." Similarly, Shane & Venkataraman (2000) argue that heterogeneity may "give rise to different entrepreneurial conjectures." Second, a more diverse set of scientists may have broader networks than those who all work in the same field (Hills et al., 1997). Networks may amplify the informational advantage mentioned above, or they may lead to sources of human or financial capital.

Our results regarding discovery team composition while holding constant the potentially confounding effect of unobserved latent commercializability of scientific advances thus help to reconcile the coordination costs of interdisciplinary work with the potential for enhanced opportunity recognition and

---

[8] One difference between the setting of prior studies vs. academic scientists is that faculty undertaking entrepreneurial ventures often remain in their positions. For example, Professor Robert Langer, whose research has spawned over 40 startups, never left MIT; instead, his associates took the lead in commercialization efforts.

resource assembly by a cross-disciplinary team. Our results on heterogeneous effects suggest that the predictive role of interdisciplinary discovery teams of academic entrepreneurship is strongest in the U.S., in the non-biotechnology life science sector, and in the pre-2000 time period. The deeper dive into the forms of interdisciplinarity that matter most likewise suggest that the coordination costs of disciplinary-focused researchers are offset by discovery-team members who are themselves interdisciplinary.

As in our prior discussion on discovery team composition, however, the interdisciplinarity results should not be interpreted as causal. In general, future research may delve more deeply into the process of scientific team formation. Boudreau et al. (2017) suggest that there are search frictions associated with the process of finding scientific collaborators. In a field experiment context, they found that randomization in research funding information session colocation among researchers had a substantial (75%) boost in the likelihood that author dyads would submit collaborative proposals. Results like these suggest that a deeper understanding of the antecedents of discovery team formation will be helpful in better understanding how discovery team composition more generally impacts academic entrepreneurship.

5. Discussion and conclusion

Given interest in the translation of academic science into commercial products, including via startups, improving our insight into the antecedents of this process is essential. Prior work has yet to account for differences in the inherent commercial potential of scientific discoveries, however, which may result in spurious inferences. Based on an algorithmic approach, we assemble a large sample (over 20,000) of scientific co-discoveries, which allows us to hold constant the scientific advance (and therefore addressing the latent commercializability confound). As compared to a population-level empirical strategy which does not address the issue of latent commercializability, we demonstrate that the magnitude and sometimes even the direction of estimated effect can differ.

Focusing on two broad classes of mechanisms—resource munificence and discovery-team composition—we confirm the importance of both in a cross-sectional analysis of all academic articles in the Web of Science from 1955-2017 while not accounting for latent commercial differences. However, when controlling for these differences via our twin-discovery approach, we no longer find empirical support for the resource munificence view. It may be that selection by researchers with a commercial disposition into prominent institutions or resource-rich geographies is a better explanation for existing findings. Regarding discovery-team composition, however, we both reaffirm and refine existing findings. In both cross-sectional and twin-discovery analyses we find a strong connection between exposure to peers with entrepreneurial experience. In controlling for latent commercialization, unlike prior literature we can state that these peer effects are not driven solely by project selection. Our results also help to reconcile two opposing forces of interdisciplinarity: although in the cross-section we find that interdisciplinary teams are less likely to commercialize—consistent with prior theory and evidence regarding coordination costs—when controlling for latent commercializability we observe that interdisciplinary teams are in fact *more* likely to commercialize.

A limitation of our methodology, however, is that we may not capture scientific commercialization by a startup that licenses or otherwise appropriates the discovery without involvement from the original scientists.[9] In addition, as previously noted, our results should also not be interpreted as causal, as team composition is of course not randomly determined. One selection effect could be the unobservability of author teams that did not successfully publish their paper. This would impact the possible censoring of observed "twin" discoveries, especially if the main reason why a given paper is not published is because journal editors decide that the focal paper is not novel given an existing paper already published or accepted for publication in the literature. If author teams of these censored papers are equally distributed by interdisciplinary and association with star commercializers, this would not present a problem. If, on the other hand, such unobserved paper author teams are much more likely to be uniform with regard to disciplinary background and less likely to have a star commercializer on the author team, then our results may be biased upwards. While we do not think this is likely, the issue illustrates a broader interpretational point associated with our methodology: we take the process generating observed scientific twins as given (and therefore exogenous to our study). As we suggest in the prior section, our findings suggest a fruitful avenue for future research would be to better understand the antecedents of discovery team composition. In addition, we believe that there is ample opportunity to examine not just outcomes related to the act of entrepreneurial commercialization itself, but also any number of subsequent venture outcomes and milestones including fundraising, job creation, and liquidity events. Doing so would inform our understanding of the fully lifecycle from scientific advance to commercial impact.

[9] This mainly applies to the PPP route of identifying commercialization, as patents can be reassigned to entities outside of the original assignee for hard-to-observe reasons (e.g., technology licensing, startup acquisition). We feel comfortable using the term "commercialization" to describe both cases, but only in the latter would we want to ascribe the commercialization to startup formation. In examining 20 PPP cases as described in Appendix B, for one or two we could not distinguish simple licensing from perhaps a chain of startup acquisitions and reassignment.

References

P. Azoulay, W. Ding, T. Stuart (2007). "The determinants of faculty patenting behavior: Demographics or opportunities?" *Journal of Economic Behavior & Organization*, 63: 599-623.

RA. Baron (2006). "Opportunity Recognition as Pattern Recognition: How Entrepreneurs 'Connect the Dots' to Identify New Business Opportunities." *Academy of Management Perspectives* 20(1):104-119.

N. Beck (2015). "Estimating grouped data models with a binary dependent variable and fixed effects: what are the issues?" *Annual Meeting of the Society for Political Methodology.*

J. Bercovitz, M. Feldman (2011). The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy*, 40: 81-93.

M. Bikard, M. Marx (2019). "Hubs as lampposts: academic location and firms' attention to science." *Forthcoming in Management Science.*

KJ Boudreau, T Brady, I Ganguli, P Gaule, E Guinan, A Hollenberg, KR Lakhani (2017). "A field experiment on search costs and the formation of scientific collaborations," *Review of Economics and Statistics*, 99(4): 565-576.

JB. Carlin, et al. (2005). "Regression models for twin studies: a critical review." *International Journal of Epidemiology*, 34.5: 1089-1099.

Cozzens, S. (1989). What do citations count? The rhetoric-first model. *Scientometrics*, 15(5-6), pp.437-447.

D. Cyranoski, N. Gilbert, H. Ledford, A. Nayar, M. Yahia. (2011). "Education: the PhD factory" *Nature*, 472: 276.

R. Fini, N. Lacetera, S. Shane (2010). "Inside or outside the IP system?" *Research Policy*, 39: 1060-1069.

J. Haltiwanger, RS. Jarmin, J. Miranda (2013). "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 95(2): 347-361.

CS. Hayter, R. Lubynsky, S. Maroulis, (2017). "Who is the academic entrepreneur? The role of graduate students in the development of university spinoffs." *Journal of Technology Transfer*, 42: 1237-1254.

G.E. Hills, G. T. Lumpkin, & R. P. Singh (1997). "Opportunity recognition: perceptions and behaviors of entrepreneurs. *Frontiers of Entrepreneurial Research*. Babson College, Wellesley MA 103-121.

DH. Hsu, T. Bernstein (1997). "Managing the university technology licensing process: findings from case studies," *Journal of the Association of University Technology Managers*, 9: 1-33.

DH. Hsu (2008). "Technology-based entrepreneurship," in S. Shane, ed., Handbook of Technology and Innovation Management, Wiley, UK.

DH. Hsu, RH. Ziedonis (2013). "Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents," *Strategic Management Journal*, 34(7): 761-781.

R. Jensen, M. Thursby (2001). "Proofs and prototypes for sale: the licensing of university inventions," *American Economic Review*, 91(1): 240-259.

M. Kenney, WR. Goe (2004). "The role of social embeddedness in professorial entrepreneurship: a comparison of electrical engineering and computer science at UC Berkeley and Stanford," *Research Policy*, 33: 691-707.

R. Kotha, G. George & K. Srikanth (2013). Bridging the mutual knowledge gap: Coordination and the commercialization of university science. *Academy of Management Journal*, 56(2), pp.498-524.

S. Lach & M. Schankerman (2008). Incentives and innovation in universities. *RAND Journal of Economics* 39(2).

E. Leahey, CM. Beckman, TL. Stanko (2017). "Prominent but less productive: The impact of interdisciplinarity on scientists' research." *Administrative Science Quarterly*, 62(1), 105-139.

J. Lerner, U. Malmendier (2013). "With a little help from my (random) friends: Success and failure in post-business school entrepreneurship," *Review of Financial Studies*, 26(10): 2411-2452.

RC. Levin, AK. Klevorick, RR. Nelson, SG. Winter (1987). "Appropriating the returns from industrial research and development." *Brookings Papers on Economic Activity*, 3:783–832.

GD. Markman, DS Siegel, M. Wright (2008). "Research and technology commercialization," *Journal of Management Studies*, 45(8): 1401-1423.

M. Marx & A. Fuegi (2020). "Reliance on Science: Worldwide front-page citations from patents to papers." *Strategic Management Journal* (forthcoming).

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

R. Nanda, JS. Sorensen (2010). "Workplace peers and entrepreneurship," *Management Science*, 56(7): 1116-1126.

RP. O'Shea, TJ. Allen, A. Chevalier, F. Roche (2005). "Entrepreneurial orientation, technology transfer and spinoff performance of U.S. universities," *Research Policy*, 34: 994-1009.

FT. Rothaermel, SD. Agung, L. Jiang (2007). "University entrepreneurship: A taxonomy of the literature," *Industrial and Corporate Change*, 16(4): 691-791.

S. Samila, O. Sorenson (2011). "Venture capital, entrepreneurship, and economic growth," *Review of Economics and Statistics*, 93(1): 338-349.

S. Shane & S. Venkataraman (2000). "The promise of entrepreneurship as a field of research," *The Academy of Management Review* 25(1): 217-226.

TE. Stuart, H. Hoang, RC Hybels (1999). "Interorganizational endorsements and the performance of entrepreneurial ventures," *Administrative Science Quarterly*, 44(2): 315-349.

T. Stuart, W. Ding (2006). "When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences," *American Journal of Sociology*, 112: 97-144.

T. Stuart, O. Sorenson (2003). "Liquidity events and the geographic distribution of entrepreneurial activity," *Administrative Science Quarterly*, 48(2): 175-201.

L. Zucker, M. Darby, M. Brewer (1998). "Intellectual human capital and the birth of US biotechnology enterprises," *American Economic Review*, 88(1): 290-305.

Table 1: Descriptive statistics and correlations for 38,715,150 articles in the Web of Science

	mean	stdev	min	max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) # authors	2.89	10.92	1	30	1.000							
(2) Author prestige	1.68	1.65	0	7.538	0.134	1.000						
(3) Institution prestige	4.90	2.52	0	10.31	0.044	0.390	1.000					
(4) Citations from industry patents	0.01	0.10	0	6.648	0.009	0.030	0.016	1.000				
(5) Interdisciplinarity of authors	0.00	0.02	0	1	0.007	0.026	0.016	0.020	1.000			
(6) Star commercializer' among authors	0.00	0.06	0	1	0.025	0.070	0.039	0.031	0.382	1.000		
(7) Prior coauthors include 'star commercializer'	0.48	0.29	0	1	-0.017	-0.053	-0.023	-0.004	-0.003	-0.008	1.000	
(8) Venture capital investments in postal code	0.40	0.75	0	5.756	-0.003	0.049	0.167	0.026	0.015	0.027	-0.010	1.000

Table 2: OLS estimates for startup-commercialization of 38,715,150 academic articles in the Web of Science, 1955-2017

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Resource Munificence								
Author prestige		0.0000546***						0.0000043***
		(0.0000016)						(0.0000016)
Institution prestige			0.0000129***					0.0000006
			(0.0000013)					(0.0000013)
Venture capital investments in postal code				0.0000833***				0.0000554***
				(0.0000059)				(0.0000059)
Discovery Team Composition								
Interdisciplinarity of authors					-0.0000384***			-0.0000157*
					(0.0000084)			(0.0000084)
'Star' commercializer among authors						0.0396470***		0.0221866***
						(0.0012537)		(0.0013032)
Prior coauthors include 'star commercializer'							0.0209219***	0.0176740***
							(0.0003505)	(0.0003506)
Controls								
# authors	0.0000029***	0.0000023***	0.0000028***	0.0000029***	0.0000029***	0.0000025***	0.0000006***	0.0000006***
	(0.0000002)	(0.0000002)	(0.0000002)	(0.0000002)	(0.0000002)	(0.0000002)	(0.0000002)	(0.0000001)
Citations from industry patents	0.0051409***	0.0051255***	0.0051388***	0.0051331***	0.0051406***	0.0049425***	0.0047368***	0.0046819***
	(0.0001491)	(0.0001491)	(0.0001491)	(0.0001490)	(0.0001491)	(0.0001478)	(0.0001467)	(0.0001464)
Constant	0.0001888***	0.0000905***	0.0001196***	0.0001548***	0.0002072***	0.0001663***	0.0001072***	0.0000812***
	(0.0000023)	(0.0000030)	(0.0000070)	(0.0000031)	(0.0000047)	(0.0000022)	(0.0000020)	(0.0000084)
R-squared	0.001	0.001	0.001	0.001	0.001	0.006	0.009	0.010

Note: All models estimated w/OLS; robust standard errors: *=p<.1; **=p<.05; ***=p<.01. Mean of the DV = 0.00019. All models have fixed effects for year of publication, country of corresponding author, and scientific field. Column (1) considers only control variables. Columns (2-7) consider controls and explanatory variables one by one. Column (8) considers controls and explanatory variables jointly.

Table 3: Descriptive statistics and correlations for 23,851 "twin" discoveries

	mean	stdev	min	p25	p50	p75	max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) # authors	6.23	4.65	1	3	5	8	30	1.000							
(2) Author prestige	2.51	1.05	0	1.79	2.48	3.18	9.5	0.674	1.000						
(3) Institution prestige	1.06	0.51	0	0.69	0.92	1.30	4.46	0.099	0.579	1.000					
(4) Citations from industry patents	0.02	0.16	0	0	0	0	4.23	0.043	0.029	-0.017	1.000				
(5) Interdisciplinarity of authors	0.46	0.27	0	0.24	0.52	0.67	0.97	0.257	0.561	0.220	0.030	1.000			
(6) Star commercializer' among authors	0.01	0.07	0	0	0	0	1	0.061	0.102	0.052	0.004	0.062	1.000		
(7) Prior coauthors include 'star commercializer'	0.03	0.17	0	0	0	0	1	0.174	0.212	0.098	0.023	0.116	0.439	1.000	
(8) Venture capital investments in postal code	0.15	0.58	0	0	0	0	6.77	0.025	0.000	-0.039	0.050	-0.006	-0.003	0.050	1.000

Table 4: Comparing commercialized vs. uncommercialized "twin" discoveries (N=23,851)

	mean		min		p25		p50		p75		max		stdev	
commercialized?	*yes*	*no*	*yes*	*no*	*yes*	*no*	*yes*	*no*	*yes*	*no*	*yes*	*no*	*yes*	*no*
# authors	8.991	6.207	1.000	1.000	3.000	3.000	7.000	5.000	12.000	8.000	30.000	30.000	7.550	4.611
Author prestige	2.816	2.508	0.000	0.000	1.609	1.792	2.708	2.485	3.714	3.178	6.841	10.012	1.379	1.058
Institution prestige	0.954	1.061	0.000	0.000	0.693	0.693	0.081	0.916	1.131	1.299	3.238	4.458	0.437	0.519
Citations from industry patents	0.032	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.302	4.234	0.207	0.152
Interdisciplinarity of authors	0.521	0.462	0.000	0.000	0.427	0.245	0.611	0.514	0.704	0.764	0.888	0.965	0.257	0.273
Star commercializer' among authors	0.094	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	0.283	0.066
Prior coauthors include 'star commercializer'	0.250	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	0.423	0.157
Venture capital investments in postal code	0.465	0.150	0.000	0.000	0.000	0.000	0.000	0.000	0.693	0.000	6.445	6.766	0.963	0.590

Table 5: OLS estimates for startup-commercialization of 23,851 "twin" discoveries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Resource Munificence								
Author prestige		0.000179						-0.00232
		(0.00136)						(0.00229)
Institution prestige			-0.00188					-0.00264
			(0.00174)					(0.00255)
Venture capital investments in postal code				0.00148				0.00181
				(0.00261)				(0.00259)
Discovery Team Composition								
Interdisciplinarity of authors					0.00680*			0.00940**
					(0.00373)			(0.00472)
'Star' commercializer among authors						0.132***		0.0993***
						(0.0330)		(0.0345)
Prior coauthors include 'star commercializer'							0.0568***	0.0388***
							(0.0119)	(0.0120)
Controls								
# authors	0.000952**	0.000925**	0.000976**	0.000954**	0.000843**	0.000813**	0.000624*	0.000867*
	(0.000396)	(0.000459)	(0.000397)	(0.000396)	(0.000393)	(0.000390)	(0.000376)	(0.000516)
Citations from industry patents	-0.000996	-0.000993	-0.00109	-0.00116	-0.000914	-0.000345	-0.000198	-0.000220
	(0.00873)	(0.00873)	(0.00873)	(0.00873)	(0.00873)	(0.00863)	(0.00884)	(0.00872)
Constant	0.00350	0.00323	0.00535*	0.00326	0.00105	0.00363	0.00395	0.00638**
	(0.00254)	(0.00320)	(0.00302)	(0.00254)	(0.00302)	(0.00253)	(0.00249)	(0.00308)
R-squared	0.520	0.520	0.520	0.520	0.520	0.525	0.524	0.527

Note: All models estimated w/OLS; robust standard errors: *=p<.1; **=p<.05; ***=p<.01. Mean of the DV= 0.0094. Each model includes fixed effects for the twin discovery. Column (1) considers only control variables. Columns (2-7) consider controls and explanatory variables one by one. Column (8) considers controls and explanatory variables jointly.

Table 6: Deeper examination of interdisciplinarity and star commercializers

	(1)	(2)	(3)	(4)	(5)
Resource Munificence					
Author prestige	-0.00232	-0.00287	-1.09e-05	-0.00146	-0.00154
	(0.00229)	(0.00219)	(0.00207)	(0.00198)	(0.00245)
Institution prestige	-0.00264	-0.000678	-0.00369	-0.00312	-0.00271
	(0.00255)	(0.00276)	(0.00254)	(0.00248)	(0.00254)
Venture capital investments in postal code	0.00181	0.00180	0.00183	0.00185	0.00202
	(0.00259)	(0.00260)	(0.00259)	(0.00259)	(0.00262)
Discovery Team Composition					
Interdisciplinarity of authors	0.00940**				0.00901*
	(0.00472)				(0.00470)
# primary scientific fields among authors		0.00402**			
		(0.00163)			
# authors who publish only in one scientific field			0.000233		
			(0.000808)		
Difference in max & mean author interdisciplinarity				-0.0141**	
				(0.00639)	
'Star' commercializer among authors	0.0993***	0.0990***	0.0992***	0.0992***	0.133***
	(0.0345)	(0.0344)	(0.0345)	(0.0345)	(0.0329)
Prior coauthors include 'star commercializer'	0.0388***	0.0387***	0.0389***	0.0393***	
	(0.0120)	(0.0120)	(0.0120)	(0.0120)	
Prior coauthors include 'star scientist' (by citations)					0.000812
					(0.00228)
Controls					
# authors	0.000867*	0.000683	0.000506	0.000583	0.000916*
	(0.000516)	(0.000489)	(0.000858)	(0.000494)	(0.000526)
Citations from industry patents	-0.000220	-0.000266	-0.000347	-0.000235	-0.000614
	(0.00872)	(0.00874)	(0.00871)	(0.00869)	(0.00863)
Constant	0.00638**	0.00426	0.00739**	0.0206***	0.00494
	(0.00308)	(0.00350)	(0.00303)	(0.00692)	(0.00323)
R-squared	0.527	0.527	0.527	0.527	0.525

Note: All models estimated w/OLS; robust standard errors: *=p<.1; **=p<.05; ***=p<.01. Mean of the DV= 0.0094. Each model includes fixed effects for the twin discovery. Column (1) matches column (8) of Table 5 to facilitate comparison. Columns (2-4) employ variations on the interdisciplinarity variable. Column (5) replaces the variable for prior collaboration with a star *commercializer* with prior collaboration with a star *scientist*.

Table 7: Heterogeneity in Commercialization by Geography, Industry, and Time Period

Sample =	All twins	Both twins in U.S.	Neither twin in U.S.	Only one twin in U.S.	Biotech-nology	non-Biotech Life Sciences	not Life Sciences	Both twins published <2000	Both twins published 2000 or later
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Resource Munificence									
Author prestige	-0.00232	-0.00492	-0.00149	-0.00149	0.000406	-0.00530	0.00194	-0.00366	-0.000493
	(0.00229)	(0.00541)	(0.00162)	(0.00368)	(0.00457)	(0.00524)	(0.00360)	(0.00227)	(0.00327)
Institution prestige	-0.00264	-0.00597	-0.00149	-0.000750	-0.00202	-0.00393	-0.00554	0.00340	-0.00677*
	(0.00255)	(0.00525)	(0.00156)	(0.00455)	(0.00610)	(0.00772)	(0.00401)	(0.00241)	(0.00388)
Venture capital investments in postal code	0.00181	-0.00120	0.00868	0.00365	0.00459	0.00276	0.00262	0.0176	0.00115
	(0.00259)	(0.00453)	(0.00566)	(0.00330)	(0.00643)	(0.00548)	(0.00354)	(0.0127)	(0.00262)
Discovery Team Composition									
Interdisciplinarity of authors	0.00940**	0.0195*	-0.00392	0.0111	-0.00371	0.0272**	0.00177	0.0128**	0.00562
	(0.00472)	(0.0104)	(0.00412)	(0.00720)	(0.00792)	(0.0112)	(0.00631)	(0.00544)	(0.00745)
'Star' commercializer among authors	0.0993***	0.151***	0.00417	0.0760	0.231***	0.170***	-0.0275	0.263***	0.0415
	(0.0345)	(0.0585)	(0.0499)	(0.0497)	(0.0896)	(0.0598)	(0.0677)	(0.0772)	(0.0362)
Prior coauthors include 'star commercializer'	0.0388***	0.0334	0.0495**	0.0377**	-0.0111	0.0676***	0.0348	0.0240	0.0432***
	(0.0120)	(0.0214)	(0.0200)	(0.0192)	(0.0204)	(0.0216)	(0.0225)	(0.0206)	(0.0141)
Controls									
# authors	0.000867*	0.00121	8.31e-06	0.00126	0.000902	0.00160	-0.000415	0.00231***	0.000251
	(0.000516)	(0.00130)	(0.000349)	(0.000828)	(0.00128)	(0.00114)	(0.000669)	(0.000708)	(0.000656)
Citations from industry patents	-0.000220	-0.00819	0.00853	0.00358	0.0632	-0.00258	0.00220	0.0125	-0.00265
	(0.00872)	(0.0138)	(0.00918)	(0.0151)	(0.0426)	(0.0136)	(0.00454)	(0.0187)	(0.00971)
Constant	0.00638**	0.0163**	0.00844***	-0.00236	0.00189	0.00335	0.00811**	-0.0101**	0.0144***
	(0.00308)	(0.00664)	(0.00296)	(0.00514)	(0.00537)	(0.00788)	(0.00392)	(0.00430)	(0.00403)
Observations	23,851	8,262	6,340	10,507	5,262	6,412	7,418	9,618	15,128
R-squared	0.527	0.550	0.509	0.497	0.514	0.524	0.519	0.546	0.524

Note: All models estimated w/OLS; robust standard errors: *=p<.1; **=p<.05; ***=p<.01. Mean of the DV= 0.0094. Each model includes fixed effects for the twin discovery. Column (1) matches column (8) of Table 5 to facilitate comparison. Columns (2-4) subsample by location of the corresponding author. Columns (5-7) subsample by scientific field: biotechnology, life sciences not including biotechnology, or non-life-sciences. Columns (8) & (9) subsample by time period.

Table 8: Robustness and Placebo Tests

DV =	commercialization via startup				randomly generated
Sample =	all twins	twins w/variation in DV	twins w/variation in DV	twins w/multiple adjacency	all twins
	(1)	(2)	(3)	(4)	(5)
Resource Munificence					
Author prestige	-0.00232	-0.209	-0.0842	-0.00368	-0.0156
	(0.00229)	(0.246)	(0.0969)	(0.00310)	(0.0108)
Institution prestige	-0.00264	-0.445	-0.179	-0.00147	0.00240
	(0.00255)	(0.370)	(0.145)	(0.00343)	(0.0143)
Venture capital investments in postal code	0.00181	0.108	0.0448	0.00204	-0.00255
	(0.00259)	(0.110)	(0.0465)	(0.00385)	(0.00795)
Discovery Team Composition					
Interdisciplinarity of authors	0.00940**	1.158*	0.471*	0.0113*	-0.00778
	(0.00472)	(0.667)	(0.262)	(0.00637)	(0.0245)
'Star' commercializer among authors	0.0993***	1.437*	0.382**	0.123***	0.0416
	(0.0345)	(0.823)	(0.183)	(0.0432)	(0.0710)
Prior coauthors include 'star commercializer'	0.0388***	1.195***	0.491***	0.0378**	-0.0250
	(0.0120)	(0.417)	(0.146)	(0.0163)	(0.0327)
Controls					
# authors	0.000867*	0.0219	0.00883	0.00170**	0.000331
	(0.000516)	(0.0340)	(0.0136)	(0.000714)	(0.00193)
Citations from industry patents	-0.000220	-0.0197	-0.0410	-0.00429	0.0323
	(0.00872)	(0.724)	(0.284)	(0.00783)	(0.0314)
Constant	0.00638**		0.474***	0.00279	0.535***
	(0.00308)		(0.147)	(0.00417)	(0.0145)
Observations	23,851	436	436	15,532	23,851
Model	OLS	cond. logit	OLS	OLS	OLS
Adjusted R-squared		0.0111			
R-squared	0.527		0.153	0.526	0.592

Note: All models include fixed effects for the twin discovery and are estimated with robust standard errors: *=p<.1; **=p<.05; ***=p<.01. Mean of the DV= 0.0094. Column (1) matches column (8) of Table 5 to facilitate comparison. Columns (2) and (3) restrict estimation to twin discoveries where only one paper in the twin-discovery pair was commercialized via startup formation. Column (4) restricts estimation to twin discoveries where we were able to establish that the paper twins were adjacently cited by multiple papers. In column (5), the DV is randomly generated.

Appendix A: Characteristics of Twin Discoveries

Our 23,851 twin discoveries range from 1973-2015 and are from more than 3,000 academic institutions in 106 countries. Figure A1 shows their temporal distribution. (There may be additional twin discoveries in the distant past, but these are hard to discover because SBIR data are available only since 1983, and patent-to-paper citations are difficult to collect pre-1976 given errors in OCR processing of patent applications. This may also explain why the modal year for a twin discover is somewhat more recent than for the entire Web of Science, 2000 vs. 1997.)

Figure A1: Temporal Distribution of Twin Discoveries

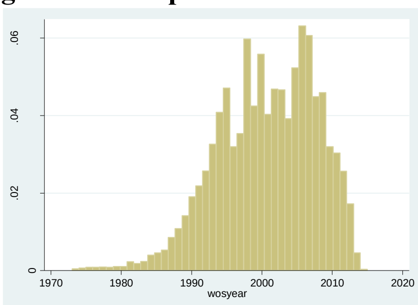

Table A1 shows the distribution of twin discoveries by geography, discipline, and institution. Over half of twin discoveries occur in the U.S., followed by Great Britain, Germany, and Japan. When considering pairs of twin papers, one-third of pairs both occur in the U.S. and 37% of twin papers are in the same country.

Panel B details the disciplinary fields of the twin discoveries. The life sciences are responsible for many of the most popular categories of twin discoveries, although Physics is the most popular category. Astronomy & Astrophysics is also a frequent source of twin discoveries. Finally, Panel C tabulates the academic institutions with the most twin discoveries.

Table A1: Twin Geography, Disciplines, and Institutions

Panel A		Panel B		Panel C	
Top 20 countries	%	Top 20 disciplines	%	Top 20 institutions	%
United States	54.1	Physics	6.0	Harvard	3.3
Great Britain	8.1	Cell Biology	5.4	UC San Francisco	1.5
Germany	6.8	Medicine, General & Internal	4.8	Stanford	1.5
Japan	5.3	Genetics & Heredity	4.0	University of Texas	1.4
France	4.5	Immunology	3.7	MIT	1.3
Canada	3.2	Astronomy & Astrophysics	2.9	UC Berkeley	1.3
Netherlands	2.1	Neurosciences	2.9	Yale	1.3
Italy	2.1	Oncology	2.6	Johns Hopkins	1.1
Switzerland	2.0	Developmental Biology	2.0	UC San Diego	1.1
Austria	1.7	Hematology	1.6	Caltech	1.0
Sweden	1.2	Physics, Condensed Matter	1.5	Columbia	0.9
China	1.1	Cardiac & Cardiovascular Systems	1.5	UCLA	0.9
Israel	0.9	Clinical Neurology	1.3	Cambridge University	0.9
Spain	0.7	Chemistry	1.2	Washington University	0.9
Denmark	0.7	Virology	1.1	University of Washington	0.9
Austria	0.6	Endocrinology & Metabolism	1.0	Tokyo University	0.9
Belgium	0.4	Geochemistry & Geophysics	1.0	University of Pennsylvania	0.8
Finland	0.4	Gastroenterology & Hepatology	0.9	University of Michigan	0.8
South Korea	0.4	Optics	0.9	Oxford University	0.8
Scotland	0.2	Chemistry, Physical	0.8	Rockefeller University	0.8

Appendix B: Name-overlap and validation for the startup commercialization outcome variable

As described in the main text, our algorithm for determining commercialization relies on overlap between the authors of a paper in the Web of Science and either inventors on a patent or the principal investigators on an SBIR award.

We implement name matching for Web of Science authors vs. SBIR personnel, removing hyphenation and other punctuation. (We examine the first 30 authors on each paper although some papers have more than 30 authors.) Although full names are available for SBIR and patents, many papers only have the authors' surname and initial(s). If both the author and the SBIR awardee have both initials present but these do not match, a score of zero is assigned. Names lacking first initials are ignored. Otherwise, a match score is assigned through a series of steps. First, we determine whether the surnames match exactly or nearly, where "nearly" indicates that both surnames are more than five characters long and fewer than ¼ of the characters must be changed to convert one to the other (i.e., Levenshtein distance). Moreover, the surnames must start with the same letter (e.g., "Rogers" and "Bogers" are not matched). Two names are treated as a preliminary match if the surname meets these criteria and the first initials also match. We want to avoid the situation where the author "J Smith" is assumed to be the same as the SBIR awardee "Jesse Smith", so we score surnames according to their inverse frequency of appearance in the Web of Science. For instance, surname Smith would be downscaled to near-zero as it is among the most common author names. Surnames that comprise less than 0.007% of all authors (i.e., 2^{nd} percentile) are not downscaled. If only two authors match between the paper and SBIR grant, and both of them represent more than 0.005% of all authors, we conclude that there is no match. Regardless of surname, matches are considered exact if both first and second initials are present for both names and they both match. A similar algorithm is implemented for computing overlap between authors of articles and inventors on patents.

To evaluate whether our algorithm truly captures instances of startup commercialization, we examine a random sample of both types of potential examples of commercialization to seek direct confirmation of our algorithmic approach. Panel A of Table B1 shows five of the 20 examples of paper-patent pairs we researched, and Panel B shows five of the 20 examples of SBIR grants. We start by randomly selecting 20 scientific papers drawn from each route of identifying commercialization. For each of these papers, we retrieve the underlying scientific article via Google Scholar searches and record the authors. For Panel A, we retrieve the associated patent from our algorithmic approach described in the main text via Google Patents (patents.google.com). We record the patent title, inventors, and assignee. For Panel B, we retrieve the associated SBIR grants to the focal companies via sbir.gov and record the grant title, funding agency and amount, and the listed principal investigator/business contact. To verify the linkages in both panels between scientific paper and commercialization activity, we conduct web searches in the following manner: we find the overlapping names between paper author and patent inventor (Panel A) or SBIR contact (Panel B) – those are shown in bold in the table. We search the web for the union of the overlapped name(s) and the new venture entity (patent assignee in Panel A; SBIR company in Panel B). The final column in both panels of the table provide web links (all accessed in January 2019) providing confirmation of commercialization activity in all ten instances (in the broader sample, we verified 39 out of 40 overall cases).

One interesting case is the second entry in Panel A. We initially had difficulty finding confirmation, but then found that one of the author/inventors, Larry Gold, had founded a company, NeXagen to commercialize his technology, changed the name of the company, and subsequently sold that company to Gilead Sciences. The patent was subsequently reassigned to Gilead Sciences, which is why initially we thought we had failed to find a linkage.

Appendix Table B1, Panel A: random sample of five patent-paper-pair instances of startup commercialization

Paper title	Journal / Year	Authors	Institution	Patent	Inventors	Patent assignee	Linkages
RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions	*PNAS* / 2011	Wiedenheft, B; van Duijin, E; Bultema, JB; Waghmare, SP; Dickman, M; Zhou, KH; Barendregt, A; Westphal, W; **Doudna, JA**	Univ Calif Berkeley	Compositions and methods of nucleic acid-targeting nucleic acids (9260752)	Andrew Paul May; **Rachel E. Haurwitz**; **Jennifer A. Doudna**; James M. Berger; Matthew Merrill Carter; Paul Donohoue	Caribou Biosciences, Inc.	**Doudna** is on Caribou's SAB; **Haurwitz** is Caribou's CEO and on the firm's BoD. Source: https://cariboubio.com/about-us
Systematic evolution of ligands by exponential enrichment - RNA ligans to bacteriophage-T4 DNA-polymerase	*Science* / 1990	Tuerk, C; **Gold, L**	Univ Colorado	Systematic evolution of ligands by exponential enrichment: tissue selex (6613526)	Joseph S. Heilig; **Larry Gold**	Gilead Sciences, Inc.	**Gold** is a founder of NeXagen, which became NeXstar Pharmaceuticas. That organization merged with Gilead Sciences in 1999. Source: https://somalogic.com/about-us/leadership/larry-gold-2/
Phase selection of microcrystalline GaN synthesized in supercritical ammonia	Journal of Crystal Growth / 2006	**Hashimoto, T**; Fujito, K; Sharma, R; **Letts, ER**; Fini, PT; Speck, JS; Nakamura, S	Univ Calif Santa Barbara	Method for producing group III-nitride wafers and group III-nitride wafers (9803293)	**Tadao Hashimoto**; Edward Letts; Masanori Ikari	SixPoint Materials Inc	**Hashimoto** is CEO/CTO of SixPoint; **Letts** is VP of Technology of the firm. Source: http://www.spmaterials.com/team.htm
Preoperative Diagnosis of Benign Thyroid Nodules with Indeterminate Cytology	*NEJM* / 2012	Alexander, EK; **Kennedy, GC;** Baloch, ZW; Cibas, ES; Friedman, L; Lanman, RB; Mandel, SJ ;Yener, N; Kloos, RT; LiVolsi, VA; Lanman, RB; Steward, DL; Friedman, L; Kloos, RT; Wilde, JI; Raab, SS; Haugen, BR; Steward, DL; Zeiger, MA; Haugen, BR	Brigham & Womens Hospital	Algorithms for disease diagnostics (9495515)	**Giulia C. Kennedy**; Darya I. Chudova; Eric T. Wang; Jonathan I. Wilde	Veracyte Inc	**Kennedy** is Chief Scientific and Medical Officer of Veracyte. https://www.veracyte.com/who-we-are/leadership/executive-team. Wilde was a director and VP of Discovery Research at Veracyte. https://uk.linkedin.com/in/jonathanwilde650
Human retinoblastoma susceptibility gene - cloning, identification, and sequence	*Science* / 1987	**Lee, WH**; Bookstein, R; Hong, F; Young, LJ; Shew, JY; Lee, EYHP	Univ Calif San Diego	Therapeutic use of the retinoblastoma susceptibility gene product (5851991)	**Wen-Hwa Lee**; Eva Y-H.P. Lee; David W. Goodrich; H. Michael Shepard; Nan Ping Wang; Duane Johnson	University of California; Canji Inc	**Wen-Hwa Lee** was Chair of the Scientific Advisory Board of Canji, Inc. http://rcndd.cmu.edu.tw/sites/default/files/WHL-CV.pdf. Canji was "formed to commercialize suppressor oncogene technology developed by Dr. Wen-Hwa Lee of the University of California at San Diego. Canji, Inc. operates as a subsidiary of Merck & Co." https://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapid=26032.

Appendix Table B1, Panel B: random sample of five SBIR instances of startup commercialization

Paper title	Journal / Year	Authors	Institution	SBIR Company	SBIR Grant(s)	SBIR PIs	Linkages
The outer mitochondrial membrane protein mitoNEET contains a novel redox-active 2Fe-2S cluster	Journal of Biological Chemistry / 2007	Wiley, SE; Paddock, ML; Abresch, EC; Gross, L; van der Geer, P; Nechushtai, R; Murphy, AN; Jennings, PA; Dixon, JE	Univ Calif San Diego	Mitokor, Inc.	"Mitochondrial Functional Proteomics" (2005 for $100,000 from the Department of Defense); "Osteoarthritis/Chondrocalcinosis: Mitochondrial Therapy" ($106,745 from the Department of Health and Human Services (HHS))	Eoin Fahy; Anne Murphy	Murphy was Director of Mitochondrial Biology at MitoKor: https://www.researchgate.net/profile/Anne_Murphy/2
Scattering theory derivation of a 3D acoustic cloaking shell	Physical Review Letters / 2008	Cummer, SA; Popa, B; Schurig, D; Smith DR; Pendry, J; Rahm, M; Starr A	Duke Univ	SensorMetrix, Inc.	"Development of Acoustic Metamaterial Applications" ($750,813 from the Dept of Defense (Navy))	Anthony Starr	Dr. Anthony Starr is the founder, president & CEO of SensorMetrix. http://www.sensormetrix.com/key-personnel.html
Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini	Cell / 2008	Mahrus, S; Trinidad, JC; Barkan, DT; Sali, A; Burlingame, AL; Wells, JA	Univ Calif San Francisco	Sunesis Pharmaceuticals, Inc.	"Development of Conformation Specific Kinase Inhibitors" (HHS for $1.5M)	James A. Wells	Wells is founder of Sunesis Pharmaceuticals. https://www.crunchbase.com/person/jim-wells#section-jobs and https://www.bloomberg.com/research/stocks/private/person.asp?personId=467474&privcapId=3768647&previousCapId=177932577&previousTitle=REZOLUTE%20INC
Curved plasma channel generation using ultraintense airy beams	Science / 2009	Polynkin, P; Kolesik, M; Moloney, JV; Siviloglou, GA; Christodoulides, DN	Univ Arizona	Nonlinear Control Strategies, Inc.	"High Power, Room Temperature 2.4- 4 micron Mid-IR Semiconductor Laser Optimization" (Department of Defense (Air Force) for $99,995 and $746,925	Jerome V Moloney	Moloney is President and corporate head of Nonlinear Control Strategies. http://www.nlcstr.com/contact.htm
Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma	Journal of Experimental Medicine / 2012	Kiel, MJ; Velusamy, T; Betz, BL; Zhao, L; Weigelin, HG; Chiang, MY; Huebner-Chan, DR; Bailey, NG; Medeiros, LJ; Bailey, NG; Elenitoba-Johnson, KSJ	Univ Michigan	Genomenon, Inc.	"Commercial Software Using High throughput Computational Techniques to Improve Genome Analysis" (HHS- National Institutes of Health, $972,083)	Mark Kiel	Kiel is a co-founder of Genomenon and Chief Science Officer. https://www.genomenon.com/about/; https://www.crunchbase.com/organization/genomenon

Appendix C: Characteristics of Star Commercializers

Appendix Table C1 provides additional information on the nature of star entrepreneurial commercializers. Only 0.4% of the more than 73 million authors in the Web of Science have had one of their discoveries commercialized by a startup. The vast majority of authors whose discoveries are commercialized by startups do so only once (mean = 1.26). Overall, less than 0.01% of all authors are ever "stars" in this respect.

Panel A of Appendix Table C1 compares stars with all other authors in the Web of Science. Perhaps unsurprisingly, stars have many more articles and citations per article, and they have been publishing longer than non-stars. Panel B details the most popular fields among stars, using 251 fields from the Web of Science. Biochemistry & Molecular Biology is the most frequent field for entrepreneurial commercialization (13.2% of all stars work primarily in this field), followed by Chemistry, Electrical & Electronic Engineering, Immunology, and Applied Physics. Panel C shows the frequency of "star" involvement in commercialized discoveries by industry and time period.

Appendix Table C1: Descriptive statistics for star entrepreneurial commercializers

Panel A: Star commercializers vs. all other authors (N: 7,164 vs. 73,923,279)

	avg. non-star	avg. star	stderr	p<
lifetime # articles	1.639	13.708	0.040	0.000
average citations per paper	13.179	30.961	0.555	0.000
# years publishing	0.899	7.423	0.035	0.000

Panel B: Most popular fields for star commercializers

Field of Study	% of stars
Biochemistry & Molecular Biology	13.2%
Chemistry, Multidisciplinary	6.5%
Engineering, Electrical & Electronic	5.1%
Immunology	4.5%
Physics, Applied	4.2%
Oncology	3.9%
Multidisciplinary Sciences	3.6%
Chemistry, Medicinal	3.6%
Cardiac & Cardiovascular Systems	3.3%
Endocrinology & Metabolism	2.9%

Panel C: Prevalence of star commercializers among commercialized discoveries

	pre-2000	2000 and later
biotech	0.27	0.22
non-biotech life sciences	0.00	0.03
non-life sciences	0.06	0.05

Appendix D: Illustrations of discovery-team composition marginal effects

Notes: Predicted probabilities of the explanatory variables related to the composition of the discovery team on a given discovery being a star commercializer are calculated from column (8) of Table 5.

Figure 1: Star commercializers on the discovery team

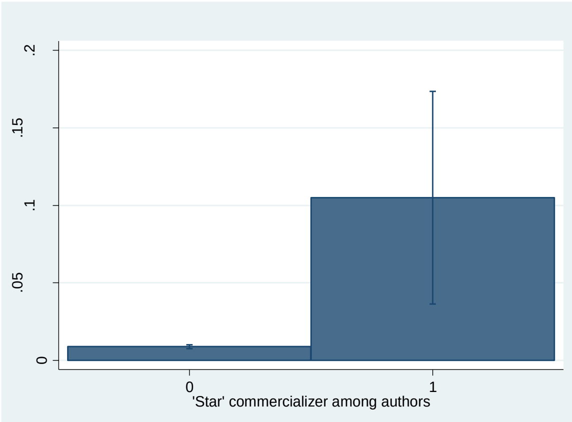

Figure 2: Prior association with star commercializers

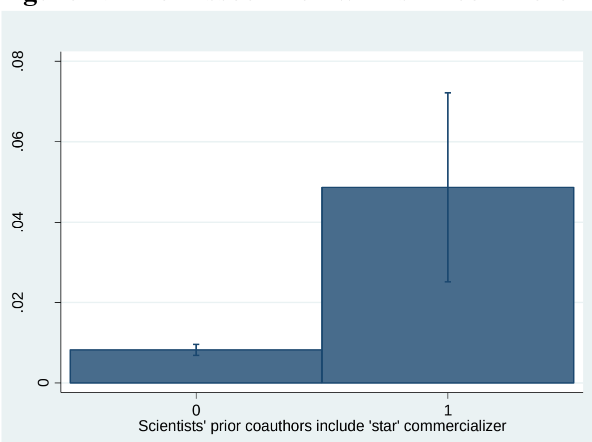

Figure 3: Predicted probabilities of startup commercialization at various levels of interdisciplinarity