Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Knowledge spillovers across funded projects and scientific career transitions

Tao Yi
North University of China
Yao Dong
dongyao33@sxu.edu.cn
Shanxi University
Jiabing Xu
Chinese University of Hong Kong
Xinyang Guo
Singapore Management University
Mingyang Sun
North University of China

Article

Keywords: public research funding, cross-project knowledge spillovers, scientific human capital, organizational learning, scientific careers

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11378062/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Knowledge spillovers across funded projects and scientific career transitions

Abstract

Public research funding is organized around discrete projects, yet knowledge generated within one project may inform scientific work beyond its original boundaries. Despite growing recognition of interdependencies among funded projects, their implications for scientists' subsequent development remain poorly understood. Drawing on organizational learning theory, we examine how cross-project knowledge spillovers relate to scientific human capital development, distinguishing knowledge circulation within research laboratories from individual exposure through prior collaboration networks. We analyze linked longitudinal panels comprising 2,281 laboratory-year and 114,820 scientist-year observations from 185 Chinese State Key Laboratories between 2012 and 2025. At the laboratory level, a one-standard-deviation increase in cross-project knowledge spillovers is associated with a 34.6% higher subsequent incidence rate of first transitions into citation-based high-impact scientific status. At the individual level, a corresponding increase in cross-boundary knowledge exposure is associated with a 24.1% higher underlying hazard of first attaining this status. These findings extend research on funding interdependencies by connecting post-award knowledge flows with the subsequent development of scientists. They also contribute to organizational learning research by distinguishing knowledge circulation across organizational activities from the relationally distributed opportunities through which individuals access that knowledge. The study highlights scientific human capital development as a relevant consideration in evaluating research funding portfolios and the knowledge environments they sustain.

***Keywords*:** public research funding; cross-project knowledge spillovers; scientific human capital; organizational learning; scientific careers

1 Introduction

In the mid-1980s, China institutionalized two approaches to organizing publicly funded research. Competitively allocated support for investigator-initiated basic research coexisted with resources directed toward national strategic and technological priorities (Chinese Academy of Sciences, 2025; National Natural Science Foundation of China [NSFC], n.d.)[1]. Balancing scientific autonomy and strategic objectives remains a longstanding research-policy challenge (Mazzucato, 2018; Wang, Lee, & Walsh, 2018). The consequences extend beyond funding allocation when differently oriented projects share organizations and scientific communities, generating knowledge valuable beyond their original boundaries. Funding arrangements can shape scientific novelty, research direction, collaboration, and subsequent careers (Wang et al., 2018; Davies, Gush, Hendy, & Jaffe, 2022; Li, Wang, Du, & Bai, 2024; Yaqub, Coburn, & Moore, 2024). China's State Key Laboratories offer an informative setting because projects supported under different funding orientations remain embedded in shared organizational and scientific environments over extended periods[2]. Although projects are the administrative units of resource allocation and evaluation, their knowledge may continue to circulate across project boundaries. Scientists may consequently develop capabilities beyond those acquired within their original funded activities, raising a question about the developmental consequences of post-award knowledge flows.

Research on funding has examined project selection and support, scientific creativity (Wang et al., 2018), and scientific risk taking (Franzoni and Stephan, 2023). Longitudinal studies connect research support with productivity, citation impact, collaboration, and careers (Graddy-Reed, Lanahan, & D'Agostino, 2021; Li et al., 2024). Participation in grant competitions can generate benefits even without receiving an award (Ayoubi, Pezzoni, and Visentin, 2019), while funding-related activities can reshape scientists' collaborative relationships (Davies et al., 2022). Meanwhile, research increasingly recognizes consequences beyond the projects initially supported. Targeted funding can generate spillovers across scientific domains (Yaqub et al., 2024), and interdependencies among projects can inform research portfolio allocation (Canton, 2025). Bozeman, Dietz, and Gaughan (2001) emphasize accumulated scientific

[1] The two approaches are represented by investigator-initiated NSFC funding and strategically oriented national programmes. Following funding reforms, several strategic programmes were later consolidated into the National Key R&D Program.

[2] State Key Laboratories host projects supported by multiple national funding schemes, allowing differently funded research to coexist within a shared organizational environment.

capabilities and relationships in evaluating research investments. Less developed is an account of how knowledge circulating among differently funded projects relates to scientists' subsequent development within a shared organization. Research on scientific agenda formation has begun to explain how team structures shape learning and spillovers (O'Kane, Mangematin, Zhang, and Haar, 2024), leaving the developmental implications of knowledge crossing funding-defined project boundaries unclear. When research organizations host projects pursuing different objectives, how does knowledge flowing among them relate to the emergence of scientists who subsequently attain high-impact scientific standing?

Organizational learning provides a theoretical basis for examining how these relationships unfold across funded research activities. A research project concentrates scientific experience around particular objectives, methods, and problem-solving trajectories, creating knowledge that may subsequently become valuable in other settings. Learning depends on how accumulated experience becomes accessible, interpreted, and incorporated into subsequent activities (Huber, 1991; March, 1991; Argote and Miron-Spektor, 2011). Inter-project learning can likewise contribute to broader organizational capabilities through the circulation and reuse of project experience (Prencipe and Tell, 2001; Brady and Davies, 2004). Within research laboratories, knowledge crossing project boundaries may expand scientists' access to concepts, methods, and scientific experience developed beyond their immediate research trajectories. Access may nevertheless vary among members. Collaborative relationships shape the accessibility and transfer of distributed knowledge (Hansen, 1999; Reagans and McEvily, 2003). Chen, Ding, Zhao, Guo, and Ning (2025) demonstrate the relevance of indirect international collaborative connections for young scientists' performance, while Xing, Ma, Fan, Sinatra, and Zeng (2025) document lasting consequences of training environments for later careers. Building on these insights, we distinguish knowledge circulation across funded activities within a laboratory from individual exposure through prior collaborators. Laboratory knowledge circulation may enlarge the experience available for scientific development across its members. Individual exposure captures how access to distributed experience varies among scientists sharing the same organization. Examining both connects organizational knowledge flows with the relational environments through which scientists encounter knowledge relevant to subsequent research.

We examine these relationships using longitudinal data linking funded projects, publications, citation flows, collaboration networks, and careers within 185 Chinese State Key Laboratories between 2012 and 2025. The organizational panel comprises 2,281 laboratory-year observations, and the individual analysis includes 114,820 scientist-year observations. Cross-project knowledge spillovers are identified through citation flows between publications associated with strategically oriented and investigator-initiated funding programmes. Laboratory-level analysis relates knowledge circulation to the subsequent incidence of scientists entering citation-based high-impact status for the first time. Individual cross-boundary knowledge exposure is measured through citation links associated with prior collaborators, excluding flows involving focal scientists' own publications. The results reveal positive relationships at both levels. A one-standard-deviation increase in laboratory spillovers is associated with a 34.6% higher subsequent incidence rate of first high-impact transitions, while a corresponding increase in individual exposure is associated with a 24.1% higher underlying first-transition hazard. The laboratory association persists across alternative measures and samples, and the individual association remains positive under alternative clustering, outcome thresholds, and weighted specifications. The findings connect cross-project knowledge circulation and differences in individual knowledge environments with subsequent scientific attainment.

Our study contributes to research on public research funding and organizational learning in two ways. First, it connects post-award project interdependence with scientific human capital development. Targeted funding can generate spillovers beyond its intended objectives, and project relationships matter for research portfolio allocation (Yaqub et al., 2024; Canton, 2025). An organizational learning perspective suggests that interdependence may also be reflected in researchers' capabilities and scientific standing within shared knowledge environments. Relating cross-project spillovers to first high-impact attainment introduces a developmental dimension to portfolio evaluation, complementing approaches focused on scientific outputs and individual funding decisions (Bozeman et al., 2001; Graddy-Reed et al., 2021). Second, the study connects organizational learning with scientific career research by distinguishing organizational knowledge circulation from knowledge accessible through individual collaborative relationships. Organizational learning emphasizes experience embedded in members and relationships (Argote and Miron-Spektor, 2011), while career research recognizes the developmental relevance of indirect connections and surrounding

environments (Chen et al., 2025; Xing et al., 2025). We bring these perspectives together by identifying cross-project exposure as a feature of prior collaboration environments and relating it to subsequent first attainment of high-impact status. The developmental relevance of funded research may therefore vary with the relational distribution of knowledge across project boundaries, informing how organizations evaluate and manage environments sustained by multiple awards.

2 Theory and hypotheses

2.1 Organizational learning theory

Organizational learning theory explains how experience generated in one setting becomes consequential for subsequent activities elsewhere (Levitt and March, 1988; Huber, 1991; Argote and Miron-Spektor, 2011). Levitt and March (1988) conceptualize organizations as learning systems in which accumulated experience shapes action through routines and established practices. Huber (1991) distinguishes knowledge acquisition, information distribution, information interpretation, and organizational memory as interconnected learning processes. Knowledge accumulated in one organizational location can therefore acquire value elsewhere when members gain access to that experience and incorporate it into subsequent activities. Argote and Miron-Spektor (2011) emphasize interactions between experience and context, with knowledge embedded in members, relationships, tasks, and organizational tools. Learning thus depends on available knowledge and the conditions under which members encounter and interpret it. March (1991) adds a dynamic dimension by distinguishing the refinement of established competencies through exploitation from the exploration of unfamiliar possibilities. Together, these perspectives suggest that organizational learning involves the circulation and subsequent use of distributed experience across organizational activities.

Publicly funded projects concentrate resources and scientific attention around identifiable objectives, participants, and periods, allowing researchers to accumulate experience along particular problem-solving trajectories (Wang et al., 2018; Davies et al., 2022). Different funding orientations may therefore generate distinct questions, methods, and expertise. Strategically oriented projects pursue national scientific and technological priorities, whereas investigator-initiated projects permit greater discretion in selecting problems. Coexistence within continuing research organizations allows their knowledge to remain scientifically connected. Inter-project learning involves articulating, codifying, and transferring experience across project boundaries

(Prencipe and Tell, 2001; Brady and Davies, 2004). Knowledge developed under one funding arrangement may consequently inform research under another, making project experience available within a broader organizational knowledge environment.

Laboratory continuity allows knowledge to acquire relevance across successive funded activities. Prencipe and Tell (2001) identify experience accumulation, articulation, and codification as processes supporting inter-project learning, while Brady and Davies (2004) connect learning within and across projects to organizational capabilities. Publications can carry findings into subsequent research, methods can reach researchers beyond their original projects, and earlier solutions can inform emerging problems. New projects also offer opportunities to reinterpret experience as scientific questions and funding priorities evolve. Cross-project knowledge flows thus connect otherwise localized research trajectories within a continuing organizational environment. Their developmental relevance extends to scientific human capital, conceptualized through scientists' accumulated knowledge, skills, and relationships (Bozeman, Dietz, and Gaughan, 2001). Experience encountered across project boundaries may expand problem-solving repertoires, introduce unfamiliar approaches, and strengthen researchers' ability to undertake consequential research. Learning can support exploration alongside the refinement of established expertise (March, 1991).

Cross-project knowledge availability does not ensure equivalent learning opportunities for laboratory members. Learning depends on how distributed experience reaches individuals and enters their activities (Huber, 1991; Argote and Miron-Spektor, 2011). Collaborative relationships provide a context for knowledge exchange because scientists differ in the research problems they pursue, the expertise they accumulate, and the knowledge sources available through collaborators. Hansen (1999) demonstrates the importance of relational connections for locating and transferring knowledge across organizational units, while Reagans and McEvily (2003) show how network relationships shape complex knowledge transfer. Scientists sharing a laboratory may consequently encounter different portions of its project experience. Cross-project circulation can enrich the organizational learning environment, while individual exposure depends partly on knowledge accessible through prior collaborators. The distinction matters when scientists encounter experience beyond their own funded activities.

The framework yields two complementary relationships between project experience and scientific development. Organizational cross-project spillovers reflect

knowledge circulation among differently funded activities and the resulting learning opportunities within laboratories. Individual cross-boundary exposure captures the extent to which scientists' prior collaboration environments connect them with knowledge originating beyond their immediate research trajectories. The relationships represent distinct dimensions of the learning environment. We therefore examine how laboratory spillovers relate to scientific talent emergence and how individual exposure through prior collaborators relates to subsequent transitions into high-impact scientific status.

2.2 Cross-project knowledge spillovers and scientific talent emergence

Scientific specialization develops deep expertise while distributing complementary knowledge across funded projects (Jones, 2009). Strategically oriented and investigator-initiated projects direct attention towards different objectives, methods, and problem-solving trajectories. Their accumulated knowledge may remain outside the activities of other scientists in the same laboratory. Cross-project spillovers connect these localized experiences when findings, methods, and insights developed under one funding orientation inform research pursued under another. Organizational learning theory suggests that knowledge becomes consequential when researchers elsewhere can interpret and use it (Huber, 1991; Argote and Miron-Spektor, 2011). Greater circulation may therefore expand the scientific experience available to laboratory members and create learning opportunities beyond their immediate funded activities.

Access to different project trajectories can enrich scientists' problem-solving repertoires. A method devised for a strategic priority may offer an unfamiliar approach to an investigator-initiated question, while exploratory findings may help reconsider a technological problem. Yaqub et al. (2024) demonstrate scientific spillovers from targeted funding beyond its original objectives, illustrating how knowledge can acquire relevance outside the agenda that generated it. March's (1991) distinction between exploration and exploitation suggests that such exposure can introduce new possibilities and refine existing expertise. Incorporating relevant experience may strengthen scientists' capabilities to identify promising questions, evaluate alternative approaches, and address problems that previously lay beyond their expertise.

An expanded repertoire may also support research with substantial scientific impact. Uzzi, Mukherjee, Stringer, and Jones (2013) show that highly cited work frequently combines conventional foundations with relatively atypical intellectual

connections. Knowledge crossing project boundaries can introduce concepts or methods unfamiliar within a scientist's established trajectory, creating opportunities to combine approaches or reinterpret evidence. Scientists who have yet to attain high-impact status may use these connections to pursue work that gains greater recognition. The opportunity is developmental because previously unfamiliar approaches can expand the range of scientific problems they are equipped to pursue. Their first transitions into that status would contribute to a higher incidence of scientific talent emergence within the laboratory. Cross-project learning thus links distributed knowledge with opportunities to develop additional high-impact researchers.

The laboratory environment can help make unfamiliar knowledge usable. Specialization raises interpretive difficulties when research problems require different expertise or methodological assumptions (Jones, 2009; Hansen, 1999). Projects within the same laboratory often share infrastructure, overlapping research communities, and accumulated experience that can help researchers recognize connections to their own problems. Prencipe and Tell (2001) emphasize experience articulation and codification in learning across projects, while Brady and Davies (2004) connect learning from individual undertakings to organizational capabilities. Successive projects provide further occasions to revisit and apply earlier knowledge as scientific questions and funding priorities evolve. Research undertaken later can therefore draw on experience that remains available after the original project's immediate objectives have been addressed. Such continuity may support capabilities that remain useful beyond the project through which researchers first encounter an approach.

Scientific and technical human capital rests on accumulated knowledge, skills, and relationships (Bozeman et al., 2001). Research funding can affect scientific careers, while funded projects can depend on one another (Graddy-Reed et al., 2021; Canton, 2025). Cross-project spillovers connect these perspectives by expanding the experience available for research and the opportunities for influential contributions. First attainment of high-impact status provides an observable milestone among scientists who have yet to achieve that standing. Stronger knowledge circulation should therefore be associated with more frequent first transitions among eligible laboratory members.

H1. *Cross-project knowledge spillovers are positively associated with the subsequent rate of scientific talent emergence within research laboratories.*

2.3 Cross-boundary knowledge exposure and individual career transitions

Knowledge circulating across funded projects may become accessible to scientists in different ways, depending on the collaborative relationships through which they encounter distributed experience. Organizational learning research emphasizes that knowledge is embedded in organizational members, relationships, and routines, making its distribution consequential for subsequent learning (Huber, 1991; Argote and Miron-Spektor, 2011). Scientists working within the same laboratory may therefore experience different learning environments, even when they share scientific infrastructure and institutional resources. Their collaborators pursue different research agendas, participate in different funded projects, and accumulate knowledge through distinct scientific activities. Collaborative relationships can consequently connect scientists to experience generated beyond their immediate project trajectories, including knowledge they have not directly contributed to producing. Cross-boundary knowledge exposure captures this individual dimension of organizational learning by reflecting the extent to which scientists' prior collaboration environments connect them with knowledge circulating between differently funded projects.

Indirect access is relevant as specialization makes identifying and evaluating knowledge across domains increasingly difficult (Jones, 2009). Hansen (1999) distinguishes searching for relevant knowledge from transferring complex experience, while Reagans and McEvily (2003) show how relational structures influence transfer. Collaborators engaging with cross-project knowledge may introduce unfamiliar findings, methods, and problem-solving experience. Repeated interaction creates opportunities to interpret that knowledge in relation to current questions. Similar collaborator counts can nevertheless conceal different knowledge environments. Some collaborators draw extensively on experience crossing project boundaries, whereas others concentrate on established trajectories. Collaboration's developmental opportunities therefore depend partly on the knowledge accessible through those relationships.

Scientific career research also directs attention to resources accessible through collaborators. Li, Aste, Caccioli, and Livan (2019) associate early collaboration with leading scientists with later academic success, while Chen et al. (2025) demonstrate benefits of indirect international connections for young scientists through knowledge and network channels. Cross-boundary exposure extends this perspective to experience generated through differently funded projects. Collaborators may introduce unfamiliar

approaches, reveal connections between research problems, or help interpret findings from other agendas. Continuing engagement provides opportunities to connect incoming knowledge with existing expertise and incorporate relevant insights into subsequent research, shaping the experience available for individual learning.

Greater exposure may strengthen scientists' capabilities to formulate questions, evaluate methods, and develop influential contributions. Accumulated knowledge, skills, and relationships support scientific development (Bozeman et al., 2001), and approaches encountered elsewhere may complement established expertise. March's (1991) distinction between exploration and exploitation suggests opportunities to pursue new directions and improve existing capabilities. Highly cited research frequently combines conventional foundations with atypical intellectual connections (Uzzi et al., 2013). Scientists whose collaborators offer greater access to cross-project knowledge may be better positioned to recognize promising combinations, reconsider assumptions, and pursue work that gains wider recognition. Across successive research activities, scientists may learn which incoming approaches complement their expertise and which combinations are worth pursuing, extending their capabilities beyond the original project.

Access to distributed knowledge may be especially consequential before first attainment of high-impact status. Research that changes a scientist's relative standing can emerge from applying accumulated expertise to promising questions or combining it with insights encountered elsewhere. Prior collaboration environments offer different opportunities for this learning within the same laboratory. First attainment provides an observable milestone for examining their developmental relevance. Greater cross-boundary knowledge exposure should therefore be associated with a higher likelihood of subsequent first attainment among scientists still eligible for the transition.

H2. *Scientists with greater prior cross-boundary knowledge exposure are more likely to attain high-impact scientific status for the first time in subsequent years.*

Figure 1 summarizes the theoretical framework.

Figure 1 near here

3 Methods

3.1 Research Design and Data

3.1.1 Analytical levels and temporal structure

We construct two linked longitudinal panels corresponding to the organizational and individual levels of our theoretical framework. The organizational panel uses the laboratory-year as the unit of analysis and examines the relationship between cross-project knowledge spillovers and the subsequent number of scientists entering high-impact scientific status for the first time. The individual panel follows scientists who remain eligible for a first high-impact transition, allowing us to examine how prior cross-boundary knowledge exposure relates to subsequent career development. Both panels are constructed from the same underlying publication, funded project, author, affiliation, and citation records.

The two panels differ in their analytical populations and eligibility requirements. Laboratory-level outcomes aggregate first transitions among scientists belonging to the eligible laboratory talent pool at the beginning of each year. The laboratory panel records 6,183 first-transition events, including 2,646 events represented in the individual-level analytical sample and 3,537 events among scientists outside that sample. The reconstructed historical individual risk set contains 236,518 scientist-year observations. Requirements concerning prior laboratory affiliation, sufficiently complete publication histories, reliable record linkage, and covariate availability yield a core sample of 114,820 scientist-year observations, which is used consistently across the four primary individual models. These differences in sample construction account for the distinct event populations examined at the two analytical levels.

We designate the outcome year as t and establish temporal ordering between prior knowledge environments and subsequent scientific attainment. Cross-project citation events, collaboration networks, prior research output and impact, project participation, external collaboration, and other time-varying covariates are measured over the three-year antecedent window from $t-3$ through $t-1$. Laboratory size and prior talent stock are measured at $t-1$, while first transitions into high-impact status are observed in year t. To construct complete antecedent windows for outcomes beginning in 2012, the underlying records extend back to at least 2009. Individual publication histories are traced further backward where necessary to determine academic age and prior high-impact status.

Publication and citation records were retrieved from the Web of Science (WoS) database on April 15, 2026, with the observation period ending on December 31, 2025. Scientific impact is assessed using field- and publication-year-normalized citation scores (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011). For each evaluation year, citation counts include only citations accumulated through December 31 of that year. Normalization benchmarks are constructed from WoS-indexed publications within the same subject category, publication year, and document type, applying the corresponding evaluation-year citation cutoff. Publications from the evaluation year are excluded from the annual impact assessment to allow sufficient citation exposure.

3.1.2 Sample coverage and disciplinary classification

Our organizational panel covers 185 Chinese State Key Laboratories between 2012 and 2025, yielding 2,281 laboratory-year observations. Individual laboratories are observed for between 7 and 14 years, with a median of 13 years, resulting in an unbalanced panel. Laboratory-years without first-transition events are retained. The individual-level core sample comprises 114,820 scientist-year observations for 23,844 scientists affiliated with the same 185 laboratories.

The original laboratory roster identifies 11 broad disciplinary categories, which we consolidate into eight groups for institutional classification and discipline-specific analyses. Agriculture is combined with biological science, manufacturing with engineering and transportation, and earth science with energy and mineral resources. Information science, medicine, chemistry, mathematical and physical sciences, and materials science remain separate. Each laboratory is assigned to its primary discipline according to the original roster. The resulting groups contain between 6 and 36 laboratories, as reported in **Supplementary Table A2**. These laboratory-level categories are used for disciplinary classification and fixed-effect specifications, while scientific impact is normalized using the finer WoS publication-field categories described in **Section 3.2.**

3.1.3 Project classification, data linkage, and sample construction

We classify funded research projects according to their dominant funding orientation. Strategically oriented projects pursue nationally defined scientific and technological priorities, whereas investigator-initiated projects generally provide researchers with greater discretion in determining research questions and approaches. The former include National Major Science and Technology Projects, the National

Science and Technology Support Program, Science and Technology Innovation 2030 Major Projects, the National Key R&D Program, the 863 and 973 Programs, and the NSFC Key Program, Major Program, and Major Research Plans. The investigator-initiated category comprises six NSFC funding schemes, including the General Program, Young Scientists Fund, Excellent Young Scientists Fund, National Science Fund for Distinguished Young Scholars, Science Fund for Creative Research Groups, and Basic Science Center Program. Classification follows the objectives and agenda-setting procedures specified in official funding documents (National Natural Science Foundation of China, 2022; State Council of the People's Republic of China, 2014; Ministry of Science and Technology of the People's Republic of China, 2016). **Supplementary Table A1** provides the detailed classification.

Publications are linked to funded projects through grant identifiers reported in funding acknowledgements and project records that explicitly attribute scientific outputs to individual awards. Laboratory affiliations are used separately to identify publications and scientists associated with State Key Laboratories. Publications acknowledging projects from both funding categories are excluded from the baseline classification, with alternative attribution rules examined in robustness analyses.

Author disambiguation combines ORCID, ResearcherID, normalized institutional affiliations, coauthor histories, research topics, and longitudinal publication trajectories (Sanyal, Bhowmick, & Das, 2021). These records allow us to reconstruct scientists' publication histories, laboratory affiliations, collaboration networks, and prior scientific attainment. For the individual analysis, scientists enter the first-transition risk set after at least two consecutive years of affiliation with the same laboratory, provided their publication histories are sufficient to establish that they have not previously attained high-impact status. They exit the risk set upon their first transition and are right-censored when they leave the laboratory or can no longer be reliably followed. Senior scientific renewal is examined using a separate analytical population. **Supplementary Table A3** summarizes the successive sample restrictions and construction of the main and matched samples.

3.2 Variable Measurement

3.2.1 High-impact status and scientific talent emergence

We measure scientists' annual research impact using the field-normalized citation impact of publications produced during the five years preceding the evaluation year. Field normalization accounts for differences in citation practices across scientific

disciplines and publication cohorts (Waltman et al., 2011). Let $P_{jt}^{(5)}$ denote the set of eligible publications produced by scientist j during years t-5 through t-1, and let NC_{pt} represent the citation impact of publication p, normalized by its field and publication year using citations accumulated through the end of year t. Annual scientific impact is calculated as

$$Impact_{jt} = \frac{1}{\left|P_{jt}^{(5)}\right|} \Sigma_{p \in P_{jt}^{(5)}} \mathrm{NC}_{pt} \tag{1}$$

We classify scientists as having high-impact scientific status when their annual research impact reaches the 90th percentile within the corresponding WoS subject category and evaluation year. Percentile-based measures are widely used to identify highly cited research while accounting for disciplinary differences in citation distributions (Waltman and Schreiber, 2013; Bornmann, 2013). The reference population comprises eligible scientists affiliated with the State Key Laboratories included in our study. Formally,

$$Top_{jt} = I\left[Impact_{jt} \geq Q_{0.90,f(j),t}\right] \tag{2}$$

where $Q_{0.90,f(j),t}$ denotes the 90th percentile of annual scientific impact within scientist j's disciplinary reference group in year t. Scientists tied at the threshold are included. Years without new publications are evaluated using previously published eligible work, while scientist-years without any eligible publications in the preceding five-year window are treated as missing.

Our primary individual-level outcome captures the first transition into high-impact scientific status. A transition occurs when a scientist enters the Top 10% group in year t without having previously attained that status during the observable publication history. We define

$$Transition_{jt} = 1\left[Top_{jt} = 1 \wedge \max_{s<t} \mathrm{Top}_{js} = 0\right] \tag{3}$$

Scientists remain in the first-transition risk set until their initial attainment of high-impact status or subsequent censoring. First attainment captures an observable milestone in scientific development, consistent with the broader scientific and technical human capital perspective emphasizing accumulated knowledge, skills, and relationships (Bozeman et al., 2001).

At the organizational level, scientific talent emergence is measured as the number of first transitions occurring among scientists belonging to a laboratory's eligible

internal talent pool at the beginning of the evaluation year. Let $M_{i,t-1}$ denote this population for laboratory i. We calculate

$$TalentEmergence_{it} = \Sigma_{j \in M_{i,t-1}} Transition_{jt} \quad (4)$$

The beginning-of-period eligible population defines the exposure underlying the laboratory-level incidence rate. Academic age is calculated as the difference between the evaluation year and the scientist's first observable publication year. Career-stage classifications and the separate senior-renewal outcome are described in **Supplementary Appendix F**.

3.2.2 Laboratory-level cross-project knowledge spillovers

Cross-project knowledge spillovers (CKS) capture citation flows between publications associated with strategically oriented and investigator-initiated research projects within the same laboratory. Citations provide observable traces of formal knowledge connections across scientific activities, although they do not encompass all forms of knowledge transmission (Aman and Gläser, 2025). We distinguish two directions of knowledge flow, from strategically oriented projects to investigator-initiated projects and in the reverse direction, denoted by $A \rightarrow B$ and $B \rightarrow A$, respectively.

The baseline measure excludes citation links for which the citing and cited publications share at least one author, thereby separating cross-project knowledge connections from continued knowledge use within the same author team. For laboratory i, outcome year t, and citation direction d, eligible cross-project citation links are calculated as

$$NetC_{d,i,t} = max\left[0, C_{d,i,t} - O_{d,i,t}\right], d \in \{A \rightarrow B, B \rightarrow A\} \quad (5)$$

where $C_{d,i,t}$ represents the number of cross-project citation links and $O_{d,i,t}$ denotes those involving author overlap. Both quantities are measured using citation events occurring during the antecedent three-year window from t-3 through t-1.

The two citation directions are subsequently aggregated, and a logarithmic transformation reduces the influence of extreme observations. Laboratory-level cross-project knowledge spillovers are defined as

$$CKS_{i,t} = ln\left[1 + NetC_{A \rightarrow B,i,t} + NetC_{B \rightarrow A,i,t}\right] \quad (6)$$

Higher values indicate greater knowledge circulation across differently funded project activities within the laboratory. **Supplementary Table C1** reports alternative directional measures, author-overlap specifications, citation-window definitions, and project-attribution restrictions.

3.2.3 Scientist-level cross-boundary knowledge exposure

Cross-boundary knowledge exposure (CBKE) captures the extent to which individual scientists are connected to cross-project knowledge flows through their prior collaborators. Collaborative relationships provide channels for accessing distributed scientific experience, while differences in collaborators' knowledge connections generate variation in the information available to individual researchers (Hansen, 1999; Reagans and McEvily, 2003).

Let $N_{j,t-1}$ denote the set of scientist *j*'s direct collaborators during the antecedent window from *t*-3 through *t*-1. For each collaborator *k*, let $L_{k,it}^{(-j)}$ represent the set of eligible cross-project citation links associated with that collaborator within laboratory *i*, covering both citation directions and excluding links involving publications authored by focal scientist *j*. Each link is uniquely identified by its citing and cited publications. Because the same citation link may be associated with multiple collaborators, we take the union of these sets to avoid duplicate counting. Individual cross-boundary knowledge exposure is measured as

$$\mathrm{CBKE}_{jit} = \ln\left[1 + \frac{\left|\bigcup_{k \in N_{j,t-1}} L_{k,it}^{(-j)}\right|}{\left|N_{j,t-1}\right|}\right], \left|N_{j,t-1}\right| > 0 \quad (7)$$

Scientist-years without prior collaborators are assigned *CBKE*=0. Normalizing the number of unique cross-project citation links by the number of prior collaborators distinguishes the intensity of surrounding knowledge exposure from collaboration-network size. Excluding citation links involving the focal scientist's publications further separates measured exposure from knowledge connections generated through the scientist's own research output.

3.2.4 Control variables

At the laboratory level, we control for laboratory size, prior talent stock, research productivity and quality, project count, funding composition, external collaboration, knowledge-base diversity, laboratory leader citation impact, and workforce composition. Laboratory size is measured as the natural logarithm of one plus the number of active researchers. Prior talent stock captures the number of scientists who had already attained high-impact status, while research quality is measured using the prior share of Top 10% publications.

Prior research productivity is calculated as the number of publications produced during the preceding three years divided by the number of active researchers at *t*-1.

$$\text{PriorProductivity}_{it} = \frac{\text{Publications}_{i,t-3:t-1}}{\text{ActiveResearchers}_{i,t-1}} \quad (8)$$

Knowledge-base diversity is measured using normalized Shannon entropy, reflecting the distribution of scientific activity across knowledge categories (Shannon, 1948; Stirling, 2007). Let p_{kit} denote the share of publications associated with knowledge category k in laboratory i, and let K represent the number of effective categories. We calculate

$$KnowledgeDiversity_{it} = -\frac{\sum_{k=1}^{K} p_{kit} \ln p_{kit}}{\ln K}, K > 1 \quad (9)$$

Categories with zero publication shares contribute zero to the summation. Project count and external collaboration scale are logarithmically transformed, while funding composition is measured as the proportion of strategically oriented projects. Workforce composition is represented by the shares of early-career and midcareer scientists, with the senior share serving as the omitted category.

The laboratory's eligible talent pool enters the organizational-level model as an exposure term, with the coefficient on $ln(TalentPool_{i,t-1})$ constrained to one (Wooldridge, 1999; Correia, Guimarães, & Zylkin, 2020). Laboratory age and time-invariant host-organization characteristics are absorbed by the laboratory and year fixed-effect structure. The continuous citation-impact measure of the laboratory leader is retained without simultaneously including the corresponding Top-status indicator.

At the scientist level, controls include academic age, prior research productivity, normalized citation impact, the share of Top 10% papers, number of collaborators, average team size, project participation, prior cross-project experience, and research-topic diversity. Laboratory size, prior talent stock, and laboratory research quality are also included to account for characteristics of the surrounding research environment. **Supplementary Table B1** provides the complete variable definitions, while descriptive statistics and correlations are reported in **Supplementary Tables B2–B5**.

3.3 Model Specification and Identification Strategy

3.3.1 Organization-level Poisson pseudo-maximum likelihood model

Scientific talent emergence is measured as the number of scientists making their first transition into high-impact scientific status within a laboratory, with the eligible population varying across laboratories and years. We therefore estimate Poisson pseudo-maximum likelihood (PPML) models that account for differences in the beginning-of-period talent pool. PPML accommodates zero-event observations and

provides consistent estimates under heteroskedasticity when the conditional mean is correctly specified (Santos Silva and Tenreyro, 2006; Correia, Guimarães, & Zylkin, 2020). The full specification is

$$\mathbb{E}(Y_{it} \mid \mathbf{X}_{it}) = \text{TalentPool}_{i,t-1}\exp\left(\beta_1 \text{CKS}_{it} + \boldsymbol{\gamma}'\mathbf{X}_{it} + \mu_i + \lambda_{f(i),t}\right) \quad (10)$$

where the dependent variable denotes the number of first high-impact transitions in laboratory i during year t, CKS captures prior cross-project knowledge spillovers, and the control vector contains pre-outcome laboratory characteristics. Laboratory fixed effects account for time-invariant organizational heterogeneity, while discipline-by-year effects absorb field-specific annual conditions. The exponentiated CKS coefficient represents the incidence-rate ratio (IRR) associated with a one-unit increase in cross-project knowledge spillovers. To facilitate interpretation, we also calculate the IRR for a one-standard-deviation increase by exponentiating the product of the coefficient and the observed CKS standard deviation. The corresponding percentage change is the resulting IRR minus one, multiplied by 100.

3.3.2 Scientist-year discrete-time hazard model

At the individual level, we examine scientists' first transitions into high-impact scientific status using a discrete-time hazard model with a complementary log-log link. The approach accommodates event times observed in annual intervals and corresponds to a grouped-time proportional-hazards specification (Prentice and Gloeckler, 1978; Allison, 1982). Let h_{jit} denote the conditional probability that scientist j, affiliated with laboratory i, makes a first high-impact transition in year t, given that the scientist remains eligible at the beginning of that year. We estimate

$$\log\left[-\log\left(1 - h_{jit}\right)\right] = \beta_1 \widetilde{\text{CBKE}}_{jit} + \boldsymbol{\gamma}'\mathbf{Z}_{jit} + g\left(\text{Age}_{jt}\right) + \mu_i + \lambda_{f(j),t} \quad (11)$$

where the standardized exposure term represents prior cross-boundary knowledge exposure, the control vector includes scientist- and laboratory-level characteristics, and the age function captures the academic-age baseline. Laboratory and discipline-by-year fixed effects account for stable laboratory characteristics and discipline-specific annual conditions.

CBKE is standardized using the estimation-sample standard deviation, allowing its exponentiated coefficient to be interpreted as the ratio of underlying hazard intensities associated with a one-standard-deviation increase in exposure. Standard errors are primarily clustered at the laboratory level using the CR1 estimator, with scientist-level clustering examined in supplementary analyses. Scientists leave the first-

transition risk set after their initial attainment, while subsequent attainment and senior renewal are analyzed separately.

3.3.3 Sensitivity analyses and temporal diagnostics

We conduct a series of sensitivity analyses to examine the stability of the estimated relationships across alternative measurements, sample restrictions, and model specifications. At the laboratory level, we distinguish the two directions of cross-project citation flows, vary the treatment of author-overlapping and prior collaborator citations, and examine alternative antecedent windows, research-quality controls, and project-attribution rules. Additional analyses exclude the pandemic years, introduce future CKS measures to examine temporal associations, and randomly reassign exposure within discipline-year groups. The laboratory specifications retain the eligible talent-pool exposure and fixed-effect structure where applicable. Detailed results are reported in **Supplementary Tables C1–C2**.

At the individual level, we examine alternative clustering, prior productivity controls, and a broader Top 20% definition of high-impact scientific status. Coarsened exact matching (CEM) provides a further comparison between scientists with relatively high and low cross-boundary knowledge exposure (Blackwell, Iacus, King, & Porro, 2009; Iacus, King, & Porro, 2012). The analysis uses the recorded exposure groups, coarsened strata, and matching weights, with strata defined by discipline, year, and career stage. The matched sample contains 36,157 scientist-year observations, of which 35,206 enter the weighted outcome model across 181 laboratories. Covariate balance is assessed using standardized mean differences calculated with a common pre-matching pooled standard deviation.

We also examine sensitivity to selective censoring using inverse probability of censoring weights (IPCW), following Robins and Finkelstein (2000) and Cole and Hernán (2008). The weighted specification retains the core outcome model and applies the supplied positive censoring weights. Scientists are censored when they leave their focal laboratory or can no longer be reliably followed, while concurrent affiliations are assigned according to the primary-laboratory rule. **Table 4** and **Supplementary Table D1** report the individual-level sensitivity analyses and matching diagnostics. Additional examinations of collaborative structures, research-topic renewal, and career-stage differences are presented in **Supplementary Appendices E–F**.

4 Results

4.1 Descriptive statistics

The laboratory panel records 6,183 first transitions across 260,658 beginning-of-period scientist-years at risk, corresponding to an annual incidence rate of 2.37%. Scientific talent emergence averages 2.711 events per laboratory-year (SD = 1.934), with 10.5% of laboratory-years recording no transitions. The mean CKS is 1.254 (SD = 0.631), ranging from 0 to 2.773, and its correlation with first-transition counts is 0.263. The variation in event counts and eligible population sizes motivates the incidence-rate specification. **Supplementary Tables B2** and **B4** present the complete laboratory-level descriptive statistics and correlations.

The individual analytical sample contains 114,820 scientist-year observations and 2,646 first transitions, yielding an annual transition incidence of 2.30%. CBKE averages 0.192 (SD = 0.207), with a range of 0 to 1.946. The 2,646 individual events are included in the laboratory total, which also comprises 3,537 transitions among scientists outside the covariate-complete individual sample. Accordingly, the laboratory analysis covers a broader eligible population, whereas the individual models examine scientists with complete information on the required covariates. Sample construction and individual-level distributions are documented in **Supplementary Tables A3**, **B3**, and **B5**.

4.2 Cross-project knowledge spillovers and laboratory talent emergence

Table 1 presents the laboratory-level PPML estimates across four specifications. Model 1 includes CKS, laboratory and year fixed effects, and the beginning-of-period TalentPool offset. Model 2 adds laboratory size, prior talent stock, productivity, and research quality; Model 3 introduces the remaining laboratory controls. Model 4 replaces common year effects with discipline-by-year effects. The CKS coefficients are 0.395, 0.455, 0.456, and 0.472, respectively, with 2,281 laboratory-year observations retained throughout. The positive association persists as observed laboratory characteristics and discipline-specific annual conditions are incorporated. With laboratory fixed effects, the estimates draw on within-laboratory changes in cross-project knowledge circulation after accounting for the included covariates and time effects.

In the most comprehensive specification, the CKS coefficient is 0.472 (SE = 0.029, $p < 0.001$), corresponding to an incidence-rate ratio (IRR) of 1.602 for a one-unit increase (95% CI = [1.514, 1.696]). Evaluated at the observed standard deviation of

CKS, the IRR is 1.346. Thus, a one-standard-deviation increase in prior cross-project knowledge spillovers is associated with a 34.6% higher expected rate of first high-impact transitions among eligible laboratory members. The finding supports H1. **Figure 2a** displays the corresponding per-standard-deviation IRRs and confidence intervals across the four models.

Table 1 near here

4.3 Cross-boundary knowledge exposure and individual first transitions

Table 2 reports the discrete-time hazard estimates for the common sample of 114,820 scientist-year observations and 2,646 first-transition events. The CBKE coefficient is 0.237 in Model 1 and 0.173 in Model 2 after controlling for prior scientist and laboratory characteristics and linear academic age. The coefficient rises to 0.216 when laboratory and discipline-by-year fixed effects are introduced in Model 3 and remains 0.216 in Model 4, which replaces linear academic age with single-year indicators. The association is therefore positive across the progressively adjusted specifications, including the model that accounts for stable laboratory differences, disciplinary time patterns, and a flexible academic-age baseline.

Model 4 yields a CBKE coefficient of 0.216 (SE = 0.024, $p < 0.001$), equivalent to a hazard ratio (HR) of 1.241 (95% CI = [1.185, 1.301]). A one-standard-deviation increase in prior cross-boundary knowledge exposure is thus associated with a 24.1% higher underlying first-transition hazard, supporting H2. The CBKE measure excludes cross-project citation links involving the focal scientist's publications and adjusts for the number of prior collaborators. The estimated association consequently captures differences in the knowledge connected to scientists through their earlier collaboration environments, conditional on the included controls. **Figure 2b** summarizes the HRs across models.

Table 2 & Figure 2 near here

4.4 Robustness and identification diagnostics

Table 3 reports the principal laboratory-level sensitivity analyses. Both directional flow measures have positive coefficients, 0.342 and 0.266, and including author-overlapping citations gives 0.463. Removing citation links associated with prior direct collaborators yields 0.414. The association thus persists when the exposure includes closer author connections and when the exclusion is tightened. Two- and four-year antecedent windows give 0.357 and 0.407, respectively. Excluding 2020–2022 gives

0.489 in 1,726 observations. **Supplementary Table C1** additionally reports the alternative MNCS quality control, dual-attribution exclusion and prior-publication-count control. These checks examine different measurement and sample choices while retaining the rate model.

The future-exposure results qualify that pattern. The two-year future-CKS coefficient is positive ($\beta = 0.187$, SE = 0.036, $p < 0.001$), whereas three- and four-year leads are negative and significant. Temporal relationships therefore extend beyond the antecedent exposure used in the main model. Common trends, anticipation or other dependence remain possible. A within-discipline-year permutation exercise yields no coefficient whose absolute value exceeds that of the observed estimate in 1,000 fits, giving a plus-one empirical p value of 0.001. Reassignment tests the observed allocation against a different benchmark. It does not resolve the significant lead associations. **Supplementary Table C2** provides the detailed temporal diagnostics.

Table 3 near here

Table 4 reports the individual-level sensitivity analyses. Scientist-clustered inference and the logarithmic prior-productivity control retain positive CBKE associations. The Top 20% definition also gives a positive coefficient of 0.227 (SE = 0.019; *HR* = 1.255), using its own 83,068-observation risk set with 3,227 events. Attaining the broader threshold removes scientists from that risk population earlier, so its sample and event count differ from the Top 10% core.

Given the supplied CEM weights, the high-versus-low exposure contrast is 0.604 (SE = 0.094; *HR* = 1.830). Several continuous covariates remain imbalanced, including productivity, prior impact, collaborator count and prior talent stock (**Supplementary Table D1**). The comparison assesses the sensitivity of the association to observed group composition. It has a binary exposure contrast and cannot be read as the same estimand as the continuous per-SD baseline. IPCW gives 0.216 (SE = 0.024) under the supplied censoring weights.

Table 4 near here

4.5 Additional analyses

To further examine the relationships between cross-project knowledge flows and scientific development, we conducted additional analyses of collaborative structures, research-topic renewal, and career-stage differences. **Supplementary Appendix E** examines collaboration breadth, depth, and organizational heterogeneity at both

analytical levels. The laboratory-level interaction estimates show limited variation in the CKS association across these collaborative dimensions. At the individual level, organizational heterogeneity exhibits a positive interaction with CBKE ($\beta = 0.051$, $p = 0.002$), suggesting that scientists' exposure to cross-project knowledge is more strongly associated with subsequent high-impact attainment in more organizationally diverse collaboration environments. The corresponding breadth and depth interactions are statistically insignificant.

Supplementary Appendix F extends the analysis to research-topic renewal and scientific development across career stages. We examine whether prior CBKE is associated with changes in research topics and whether topic shifts precede subsequent first transitions. The estimated relationships are statistically insignificant. Meanwhile, laboratory-level analyses reveal positive CKS associations with first high-impact transitions among early-career, midcareer, and senior scientists, as well as renewed high-impact attainment among senior scientists (IRRs = 1.540, 1.603, 1.578, and 1.434, respectively). The pooled individual analysis further examines career-stage variation, with the joint interaction test yielding $p = 0.209$. These supplementary analyses extend the empirical assessment of cross-project knowledge flows to different collaborative environments, dimensions of scientific search, and stages of scientific careers.

5 Discussion

5.1 Theoretical implications

Public research funding is typically administered through individual awards, although the knowledge generated under one award can acquire value in subsequent scientific activity elsewhere. Research on targeted funding documents spillovers beyond the intended research domain (Yaqub et al., 2024), and Canton's (2025) portfolio perspective recognizes interdependencies among projects when resources are allocated. Our findings extend this conversation to the development of scientists within research organizations after allocation. Laboratories exhibiting greater cross-project knowledge circulation also experience higher subsequent rates of first attainment of high-impact scientific status. Inter-project learning research explains how experience developed within a bounded undertaking can be articulated, retained, and brought into later work (Prencipe and Tell, 2001; Brady and Davies, 2004). Within funded science, such continuity makes project interdependence relevant to the researchers whose capabilities develop through successive scientific activities. The laboratory thus

becomes a setting in which the developmental relevance of connections among administratively distinct awards can be examined.

Scientific and technical human capital provides a further perspective on the value generated by these connections. Bozeman, Dietz, and Gaughan (2001) locate accumulated expertise and relationships within the outcomes of research activity, inviting evaluation beyond outputs attributed to individual grants. First entry into high-impact status captures one observable developmental milestone within this broader conception. It identifies emerging attainment among scientists who had not previously reached that standing, and the laboratory-level incidence relates those transitions to the population still eligible to experience them. The organizational finding consequently places the emergence of additional high-impact researchers alongside scientific outputs in the study of funded project portfolios. It raises a longitudinal question for research evaluation about who develops within the knowledge environment sustained by multiple awards.

A second contribution concerns the distribution of learning opportunities among scientists working within the same organization. Huber (1991) and Argote and Miron-Spektor (2011) emphasize how organizational experience becomes embedded in members and relationships and gains relevance through its circulation and interpretation. Our individual-level analysis gives relational access a more specific empirical meaning by examining the cross-project citation links associated with scientists' prior collaborators. Excluding the focal scientist's authored publications and accounting for collaborator count direct attention toward the knowledge connected through those relationships. The positive association with first high-impact attainment complements research linking coauthorship, research environments, and indirect ties with scientific careers (Li et al., 2019; Graddy-Reed et al., 2021; Chen et al., 2025). It also suggests why the existence of a collaboration tie tells only part of the story. Scientists embedded in the same laboratory may encounter different bodies of project knowledge through their respective collaborators. Distinguishing organizational circulation from individual access therefore connects organizational learning to scientific career development while preserving the different populations and outcomes represented by the two analyses.

5.2 Implications for research funding and management

The findings suggest that assessment of funded research portfolios could incorporate the development of scientists over time alongside outputs assigned to

individual awards. Indicators of first high-impact attainment, persistence in scientific performance, and participation across projects would complement established research evaluation practices informed by scientific and technical human capital (Bozeman et al., 2001). Tracking new attainment among scientists who remain eligible to achieve it could also reveal whether a laboratory's research environment is associated with the emergence of additional high-impact researchers. Such measures would preserve grant-level accountability while making the developmental dimension of project interdependence more visible.

Portfolio management may also benefit from considering which project connections make scientifically relevant experience accessible. Funding arrangements can shape creativity and collaboration (Wang et al., 2018; Davies et al., 2022). Laboratory managers might therefore experiment with problem-focused seminars, shared methodological facilities, or exchanges between teams whose research questions exhibit meaningful complementarity. In the meantime, evaluating these activities would require attention to the relevance of knowledge encountered, participants' subsequent engagement, and the coordination costs involved (Cummings and Kiesler, 2005).

An individual perspective adds the question of who can reach the knowledge circulating within a laboratory. Collaboration counts alone may conceal differences in the project knowledge linked to particular scientists. Mapping opportunities for access, including contacts across organizational settings, could inform the design and evaluation of targeted exchanges.

5.3 Limitations and future research

The findings also suggest several directions for future research. First, our longitudinal design establishes temporal ordering between prior knowledge exposure and subsequent scientific attainment, although significant associations involving future CKS leave some temporal dependence unresolved. Future studies could exploit funding policy changes, eligibility discontinuities, or evaluated cross-project exchange interventions to provide stronger evidence on the developmental consequences of knowledge spillovers.

Second, citation-based measures capture selected dimensions of organizational learning and scientific development. Cross-project citations provide observable traces of codified knowledge connections, while informal exchanges and the actual use of knowledge may remain unobserved (Aman and Gläser, 2025). Likewise, first

attainment of high-impact scientific status represents one dimension of the broader scientific human capital construct (Bozeman et al., 2001). Future research could combine bibliometric records with interviews, knowledge reuse indicators, and alternative measures of scientific achievement to better understand how cross-project learning contributes to scientists' developing capabilities.

Finally, the institutional characteristics of Chinese State Key Laboratories provide a distinctive setting for examining interactions among differently funded research projects. However, funding arrangements and organizational learning environments may vary across institutional and national contexts. Comparative studies involving universities, other public research organizations, and corporate laboratories could further clarify how project characteristics, collaborative relationships, and institutional arrangements shape the developmental relevance of cross-project knowledge flows.

5.4 Conclusion

Knowledge generated through funded projects can remain relevant as scientists pursue subsequent research. Across Chinese State Key Laboratories, cross-project spillovers are positively associated with the laboratory rate of first attainment of high-impact status, and exposure through prior collaborators is positively associated with the individual first-transition hazard. The two relationships connect post-award interdependence with scientific human capital development while distinguishing organizational knowledge circulation from individual access. Their relevance for funding evaluation lies in extending attention from attributed outputs towards the scientists developing within a shared research environment. The collaboration and topic analyses qualify how those opportunities might become consequential. Future research can examine the relevance of encountered knowledge, its interpretation and its incorporation into scientific practice. Identifying these processes would help explain when access supports durable capabilities and how research organizations can create effective opportunities for exchange.

Data Availability

The prepared laboratory-year and scientist-year analysis datasets, variable codebooks, statistical code and reproduction instructions are available for peer review through the Harvard Dataverse preview link at https://dataverse.harvard.edu/previewurl.xhtml?token=b1e0e5a1-a1f7-435d-b97c-609516da0ff6. The named version and linkage mapping can be provided upon a request stating a reasonable justification, subject to approval by the corresponding author.

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. Sociological Methodology, 13, 61–98. https://doi.org/10.2307/270718

Aman, V., and Gläser, J. (2025). Investigating knowledge flows in scientific communities: The potential of bibliometric methods. Minerva, 63(1), 155–182. https://doi.org/10.1007/s11024-024-09542-2

Argote, L., and Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. Organization Science, 22(5), 1123–1137. https://doi.org/10.1287/orsc.1100.0621

Ayoubi, C., Pezzoni, M., and Visentin, F. (2019). The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions? Research Policy, 48(1), 84–97. https://doi.org/10.1016/j.respol.2018.07.021

Blackwell, M., Iacus, S., King, G., and Porro, G. (2009). cem: Coarsened exact matching in Stata. The Stata Journal, 9(4), 524–546. https://doi.org/10.1177/1536867X0900900402

Bornmann, L. (2013). How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes, and top-cited papers. Journal of the American Society for Information Science and Technology, 64(3), 587–595. https://doi.org/10.1002/asi.22792

Bozeman, B., Dietz, J. S., and Gaughan, M. (2001). Scientific and technical human capital: an alternative model for research evaluation. International Journal of Technology Management, 22(7–8), 716–740. https://doi.org/10.1504/ijtm.2001.002988

Brady, T., and Davies, A. (2004). Building project capabilities: From exploratory to exploitative learning. Organization Studies, 25(9), 1601–1621. https://doi.org/10.1177/0170840604048002

Canton, E. (2025). A portfolio approach to research funding. Research Policy, 54(1), 105129. https://doi.org/10.1016/j.respol.2024.105129

Chen, K., Ding, Y., Zhao, B., Guo, R., and Ning, L. (2025). Benefits beyond the local network: Does indirect international collaboration ties contribute to research performance for young scientists? Research Policy, 54(5), 105233. https://doi.org/10.1016/j.respol.2025.105233

Chinese Academy of Sciences. (2025). Building a scientific monument to New China through patriotic dedication [In Chinese]. https://www.cas.cn/xzfc/202502/t20250228_5048294.shtml (accessed October 9, 2026).

Cole, S. R., and Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. American Journal of Epidemiology, 168(6), 656–664. https://doi.org/10.1093/aje/kwn164

Correia, S., Guimarães, P., and Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. The Stata Journal, 20(1), 95–115. https://doi.org/10.1177/1536867X20909691

Cummings, J. N., and Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. Social Studies of Science, 35(5), 703–722. https://doi.org/10.1177/0306312705055535

Davies, B., Gush, J., Hendy, S. C., and Jaffe, A. B. (2022). Research funding and collaboration. Research Policy, 51(2), 104421. https://doi.org/10.1016/j.respol.2021.104421

Franzoni, C., and Stephan, P. (2023). Uncertainty and risk-taking in science: Meaning, measurement and management in peer review of research proposals. Research Policy, 52(3), 104706. https://doi.org/10.1016/j.respol.2022.104706

Graddy-Reed, A., Lanahan, L., and D'Agostino, J. (2021). Training across the academy: The impact of R&D funding on graduate students. Research Policy, 50(5), 104224. https://doi.org/10.1016/j.respol.2021.104224

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. Administrative Science Quarterly, 44(1), 82–111. https://doi.org/10.2307/2667032

Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. Organization Science, 2(1), 88–115. https://doi.org/10.1287/orsc.2.1.88

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1–24. https://doi.org/10.1093/pan/mpr013

Jones, B. F. (2009). The burden of knowledge and the “death of the Renaissance man”: Is innovation getting harder? Review of Economic Studies, 76(1), 283–317. https://doi.org/10.1111/j.1467-937x.2008.00531.x

Levitt, B., and March, J. G. (1988). Organizational learning. Annual Review of Sociology, 14(1), 319–338. https://doi.org/10.1146/annurev.so.14.080188.001535

Li, M., Wang, Y., Du, H., and Bai, A. (2024). Motivating innovation: The impact of prestigious talent funding on junior scientists. Research Policy, 53(9), 105081. https://doi.org/10.1016/j.respol.2024.105081

Li, W., Aste, T., Caccioli, F., and Livan, G. (2019). Early coauthorship with top scientists predicts success in academic careers. Nature Communications, 10(1), 5170. https://doi.org/10.1038/s41467-019-13130-4

March, J. G. (1991). Exploration and exploitation in organizational learning. Organization Science, 2(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

Mazzucato, M. (2018). Mission-oriented innovation policies: challenges and opportunities. Industrial and Corporate Change, 27(5), 803–815. https://doi.org/10.1093/icc/dty034

Ministry of Science and Technology of the People's Republic of China. (2016). The significance of launching the National Key R&D Program. https://www.most.gov.cn/xwzx/twzb/gjzdyfjh/twzbbjcl/201602/t20160216_124112.html (accessed September 10, 2026).

National Natural Science Foundation of China. (2022). National Natural Science Fund Guide to Programs 2022. https://www.nsfc.gov.cn/english/site_1/pdf/NationalNaturalScienceFundGuidetoPrograms2022.pdf (accessed September 10, 2026).

National Natural Science Foundation of China. (n.d.). NSFC introduction. https://www.nsfc.gov.cn/english/site_1/index.html (accessed September 10, 2026).

O'Kane, C., Mangematin, V., Zhang, J. A., and Haar, J. (2024). How research agendas are framed: Insights for leadership, learning and spillover in science teams. Research Policy, 53(7), 105029. https://doi.org/10.1016/j.respol.2024.105029

Prencipe, A., and Tell, F. (2001). Inter-project learning: processes and outcomes of knowledge codification in project-based firms. Research Policy, 30(9), 1373–1394. https://doi.org/10.1016/s0048-7333(01)00157-3

Prentice, R. L., and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. Biometrics, 34(1), 57–67. https://doi.org/10.2307/2529588

Reagans, R., and McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. Administrative Science Quarterly, 48(2), 240–267. https://doi.org/10.2307/3556658

Robins, J. M., and Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics, 56(3), 779–788. https://doi.org/10.1111/j.0006-341X.2000.00779.x

Santos Silva, J. M. C., and Tenreyro, S. (2006). The log of gravity. The Review of Economics and Statistics, 88(4), 641–658. https://doi.org/10.1162/rest.88.4.641

Sanyal, D. K., Bhowmick, P. K., and Das, P. P. (2021). A review of author name disambiguation techniques for the PubMed bibliographic database. Journal of Information Science, 47(2), 227–254. https://doi.org/10.1177/0165551519888605

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

State Council of the People's Republic of China. (2014). Plan for deepening reform of the management of centrally financed science and technology programs, special projects and funds (State Council Document No. 64 of 2014). https://www.moe.gov.cn/s78/A05/gongzuo/s7656_zyczky/201702/t20170214_296217.html (accessed September 10, 2026).

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. Journal of the Royal Society Interface, 4(15), 707–719. https://doi.org/10.1098/rsif.2007.0213

Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. (2013). Atypical combinations and scientific impact. Science, 342(6157), 468–472. https://doi.org/10.1126/science.1240474

Waltman, L., and Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. Journal of the American Society for Information Science and Technology, 64(2), 372–379. https://doi.org/10.1002/asi.22775

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: An empirical analysis. Scientometrics, 87(3), 467–481. https://doi.org/10.1007/s11192-011-0354-5

Wang, J., Lee, Y. N., and Walsh, J. P. (2018). Funding model and creativity in science: Competitive versus block funding and status contingency effects. Research Policy, 47(6), 1070–1083. https://doi.org/10.1016/j.respol.2018.03.014

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. Journal of Econometrics, 90(1), 77–97. https://doi.org/10.1016/S0304-4076(98)00033-5

Xing, Y., Ma, Y., Fan, Y., Sinatra, R., and Zeng, A. (2025). Academic mentees thrive in big groups, but survive in small groups. Nature Human Behaviour, 9(5), 902–916. https://doi.org/10.1038/s41562-025-02114-8

Yaqub, O., Coburn, J., and Moore, D. A. Q. (2024). Research-targeting, spillovers, and the direction of science: Evidence from HIV research-funding. Research Policy, 53(8), 105076. https://doi.org/10.1016/j.respol.2024.105076

Table 1.

Laboratory-level PPML estimates of cross-project knowledge spillovers and scientific talent emergence

Variable	Scientific talent emergence			
	Model 1	Model 2	Model 3	Model 4
Cross-project knowledge spillovers	0.395***(0.024)	0.455***(0.028)	0.456***(0.028)	0.472***(0.029)
Laboratory size	—	-0.110(0.096)	-0.095(0.102)	-0.116(0.104)
Prior talent stock	—	-0.018***(0.004)	-0.018***(0.004)	-0.019***(0.004)
Prior per-capita productivity	—	-0.045(0.114)	-0.050(0.114)	-0.043(0.116)
Prior research quality	—	0.588(1.178)	0.572(1.181)	0.514(1.196)
Project count	—	—	-0.069(0.064)	-0.078(0.066)
Mission-driven project share	—	—	0.155(0.098)	0.145(0.101)
External collaboration scale	—	—	-0.078(0.066)	-0.068(0.066)
Knowledge-base diversity	—	—	3.108(7.009)	3.719(7.584)
Leader citation impact	—	—	-0.009(0.015)	-0.011(0.016)
Early-career scientist share	—	—	0.147(0.987)	0.414(0.989)
Midcareer scientist share	—	—	0.290(0.697)	0.472(0.727)
Laboratory fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	No
Discipline × year fixed effects	No	No	No	Yes
Observations	2,281	2,281	2,281	2,281

Notes. The PPML offset is ln(TalentPool), with its coefficient fixed to one. Laboratory-clustered CR1 standard errors use the independent design rank, including fixed effects. Two-sided normal-Wald p values and confidence intervals are computed from unrounded estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.

Scientist-level discrete-time hazard models of CBKE and first career transition

Variable	First career transition			
	Model 1	Model 2	Model 3	Model 4
CBKE	0.237***(0.017)	0.173***(0.027)	0.216***(0.024)	0.216***(0.024)
Individual career-background controls	No	Yes	Yes	Yes
Laboratory-level controls	No	Yes	Yes	Yes
Laboratory fixed effects	No	No	Yes	Yes
Discipline × year fixed effects	No	No	Yes	Yes
Flexible academic-age baseline hazard	No	No	No	Yes
HR	1.267	1.189	1.241	1.241
Approx. 95% CI for HR	[1.227, 1.309]	[1.128, 1.253]	[1.184, 1.300]	[1.185, 1.301]
Scientist-year observations	114,820	114,820	114,820	114,820

Notes. All columns use 114,820 core observations and 2,646 events. CBKE is standardized with the sample SD. Model 2 adds controls and linear academic age, Model 3 adds laboratory and discipline-by-year effects, and Model 4 uses single-year age indicators. Standard errors are laboratory-clustered CR1. Individual controls are prior productivity, impact, Top 10% paper share, collaborator count, team size, project participation, cross-project experience and topic diversity. Laboratory controls are size, prior talent stock and research quality. HR intervals use unrounded coefficients and covariance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.

Laboratory robustness and temporal diagnostics

Specification	***β***	***SE***	***p***	***N***
Baseline	0.472	0.029	<0.001	2,281
A→B flows	0.342	0.030	<0.001	2,281
B→A flows	0.266	0.029	<0.001	2,281
Include author overlap	0.463	0.030	<0.001	2,281
Exclude prior direct collaborators	0.414	0.029	<0.001	2,281
Two-year antecedent window	0.357	0.027	<0.001	2,281
Four-year antecedent window	0.407	0.037	<0.001	2,281
Exclude 2020–2022	0.489	0.035	<0.001	1,726
Future CKS two-year lead	0.187	0.036	<0.001	1,911
Future CKS three-year lead	-0.159	0.038	<0.001	1,726
Future CKS four-year lead	-0.176	0.036	<0.001	1,541
Within-discipline-year permutation	—	—	0.001	1,000 fits

Notes. Entries are PPML log-rate coefficients with laboratory-clustered CR1 standard errors and the ln(TalentPool) offset. A denotes strategically oriented projects and B investigator-initiated projects. Regression models include laboratory and discipline-by-year effects and the full laboratory controls. Future-exposure windows are bounded by observed calendar support through 2025. The permutation row reports the plus-one empirical p value from 1,000 successful within-discipline-year reassignments. Full specifications appear in Supplementary Tables C1–C2.

Table 4.

Sensitivity analyses for scientist-year first-transition hazards

Scientist-level specification	Coefficient or statistic	SE	HR or interpretation	p value	N
Top 10% talent threshold	0.216***	0.024	HR = 1.241	<0.001	114,820
Top 20% talent threshold	0.227***	0.019	HR = 1.255	<0.001	83,068
CEM-weighted high-versus-low comparison	0.604***	0.094	HR = 1.830	<0.001	35,206
IPCW	0.216***	0.024	HR = 1.241	<0.001	114,820
Scientist-clustered standard errors	0.216***	0.021	HR = 1.241	<0.001	114,820
ln(1+prior productivity) control	0.217***	0.024	HR = 1.242	<0.001	114,820

Notes. Continuous CBKE is standardized within each estimation sample. CEM instead contrasts supplied high- and low-exposure groups using the saved estimation flag and weights. IPCW uses the supplied censoring weights. Baseline controls, laboratory and discipline-by-year effects and single-year age indicators are retained. Standard errors are laboratory-clustered CR1 except in the scientist-clustered row. Top 20% uses its separate risk set and 3,227 events. Supplementary Table D1 reports covariate balance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1.

Knowledge flows and scientific human capital development at two analytical levels

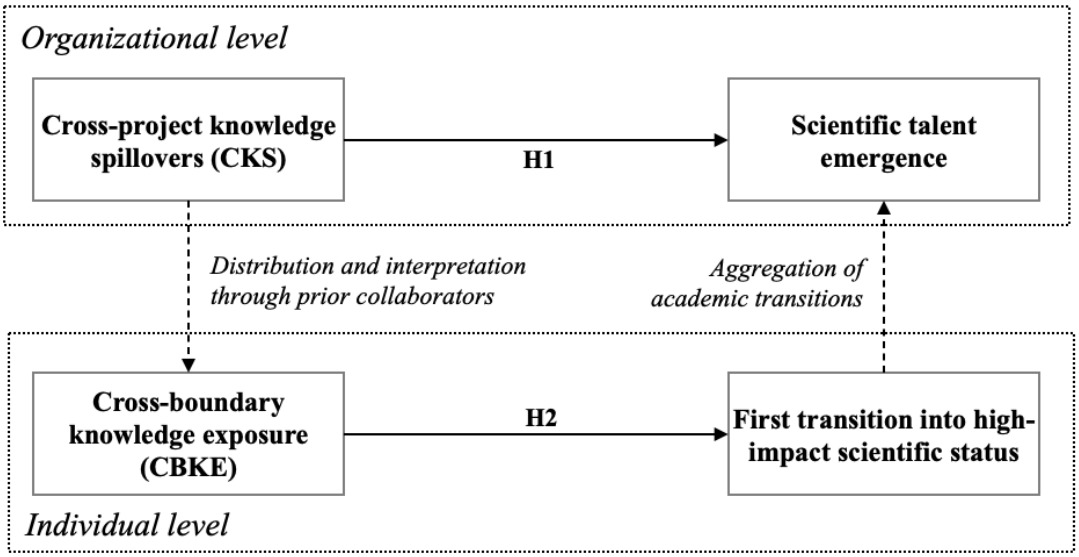
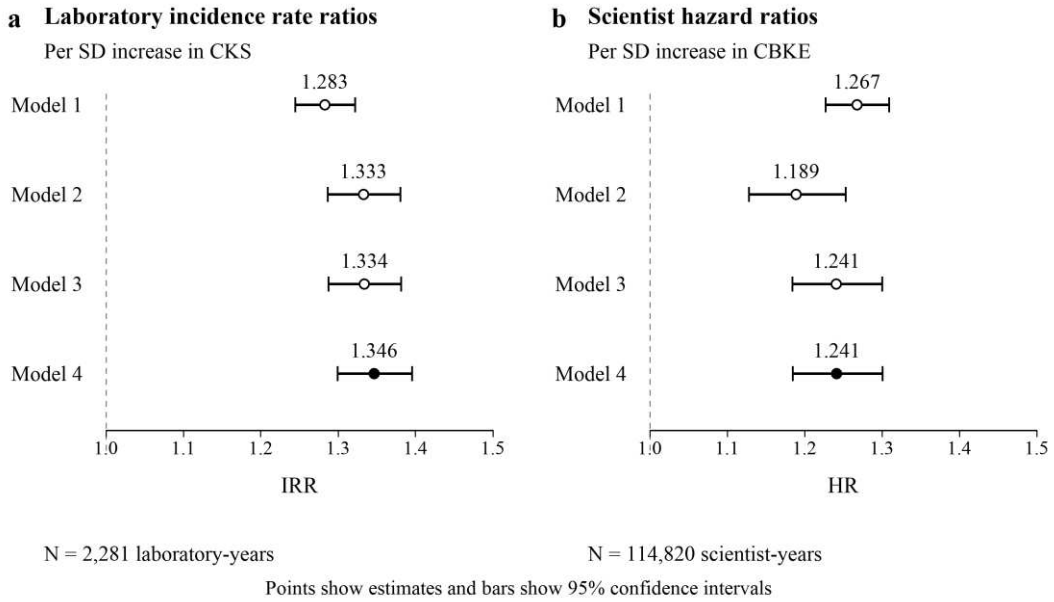


Figure 2.

Baseline associations across progressively adjusted specifications



Notes. Panel a transforms the CKS coefficients and laboratory-clustered 95% confidence intervals from Table 1 using the common-sample CKS standard deviation. Panel b exponentiates the standardized CBKE coefficients and corresponding intervals from Table 2. Models 1–4 follow their respective table definitions. All laboratory columns use 2,281 observations and all scientist columns use 114,820 observations. IRRs and HRs concern distinct outcomes, measures and populations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials.docx