www.sciencemag.org/content/354/6312/aaf5239/suppl/DC1

Supplementary Materials for

Quantifying the evolution of individual scientific impact

Roberta Sinatra, Dashun Wang, Pierre Deville, Chaoming Song, Albert-László Barabási*

*Corresponding author. Email: alb@neu.edu

Published 4 November 2016, *Science* **354**, aaf5239 (2016)
DOI: 10.1126/science.aaf5239

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S49
References

Other Supplementary Materials for this manuscript include the following:
(available at www.sciencemag.org/content/354/6312/aaf5239/suppl/DC1)

Supplementary data as zipped archive (plain text and readme files)

Contents

S1 Data Description **4**
- S1.1 APS data . . . 4
- S1.2 Web of Science and Google Scholar data . . . 6
- S1.3 Dataset selection . . . 7
- S1.4 Choosing different subsets of scientists in the APS dataset . . . 8
- S1.5 Equivalence of APS and WoS data to study physicists . . . 9
- S1.6 Publication impact . . . 9
- S1.7 Excluding review papers . . . 10

S2 Impact patterns preceding and following the highest impact paper **10**
- S2.1 Smoothing techniques . . . 10
- S2.2 Detection of trends preceding and following the highest impact paper . . . 11

S3 Random Impact rule **12**
- S3.1 Career patterns and the random impact rule in the literature . . . 12
- S3.2 Random impact rule in the WoS dataset . . . 12
- S3.3 Random impact rule for different subsets of APS scientists . . . 12
- S3.4 Impact autocorrelation . . . 13

S4 Models **13**
- S4.1 Fitting the impact distribution $P(c_{10})$ . . . 13
- S4.2 R-model and its predictions . . . 14
- S4.3 The Q-model . . . 15
- S4.4 Maximum-Likelihood estimation of the Q-model parameters . . . 16
- S4.5 Computing a scientist's Q parameter . . . 19
- S4.6 Goodness of the Q-model . . . 20
- S4.7 Q-model and the real data-generating process . . . 21
- S4.8 Q-model for the WoS dataset . . . 21
- S4.9 Stability and measurement accuracy of the Q-parameter . . . 22
- S4.10 Predictions of the Q-model: highest impact paper . . . 23
- S4.11 Predictions of the Q model: impact indicators . . . 24
- S4.12 Predictive Power of the Q-parameter . . . 26

S5 Relation between the Q-parameter and productivity **28**

S6 Controlling for coauthorship effects **28**
- S6.1 Effect on the randomness of the highest impact paper . . . 29
- S6.2 Effect on the Q parameter . . . 30

S7 Receiving Operating Characteristic (ROC) curve **31**
S7.1 Early prediction of Nobel Laureates 31

S8 Supplementary Figures **33**
S8.1 Characterization of the datasets 33
S8.2 Impact trends before and after the highest impact work 37
S8.3 Random Impact Rule 42
S8.4 Impact distribution 53
S8.5 Goodness of the Q-model and stability of the Q-parameter 54
S8.6 Predictive power of the Q-parameter 59
S8.7 Robustness of the core results for different dataset selections 62
S8.8 Robustness of the core results when controlling for coauthorship effects 66
S8.9 Prediction of independent recognitions 71
S8.10 Early impact vs productivity 74

S1 Data Description

This work is based on the analysis and modelling of two datasets of individual careers: one obtained from APS data, the other using a combination of Google Scholar and Web of Science data. The empirical findings and model predictions are qualitatively similar for all datasets. The results showed in the manuscript refer to the first dataset, while we report in dedicated sections of this Supplementary Materials the results for the second dataset.

S1.1 APS data

The dataset provided by the American Physical Society (APS) (*46*) consists of all the papers published in Physical Review, spanning across 9 different journals: Physical Review A, B, C, D, E, I, L, ST and Review of Modern Physics, from 1893 to 2010, amounting to over $450,000$ publications. For each paper the data set includes title, date of publication (day, month, year), names and affiliations of every author, and a list of the previous Physical Review papers cited. Since the datasets contains only citations between Physical Review papers, a paper's number of citations in this dataset is deflated in respect to the number reported in other databases, like Web of Science or Google Scholar.

Of all the publications, we consider only those for which: (i) there is no ambiguity between an author and his/her affiliation (an ambiguity is present when more than one author and more than one affiliation are given without any link between them); (ii) there are no more than 10 different authors. Criterion (i) is necessary to associate at least one affiliation to each author, which is a crucial step for the author name disambiguation (see section S1.1); criterion (ii) is necessary to identify those publications where each author can be considered to have a substantial contribution. The threshold of 10 authors has been chosen after the inspection of the distribution of number of authors per paper (Fig. S1). The probability density function can be approximated by a power-law, in line with previous studies (*47*). We observe a deviation from the power-law for papers containing more than 10 authors, suggesting different retribution characterizing these large collaborations *(31, 48-50)*. The application of criteria (i) and (ii) gives us a final set of $425,369$ publications.

Author name disambiguation in the APS data

The APS does not maintain unique author identifiers. Therefore, to associate a scientist to all his/her publications we need to infer first each author's actual identity. Attributing to just one scientist all publications reporting the same author name is not a valid approach. Indeed, two distinct authors may have the same full name, or the same initials and same last name; also, it happens that the same author may have slightly different names in different publications, since information about first names and middle names is sometimes incomplete or missing.

To this end, we conducted a comprehensive disambiguation process, and we describe below a hierarchical disambiguation procedure that overcomes these problems and infers actual

author identity based not only on the author name, but also on the metadata available for each publication (*47*). First, we consider each author of each publication to be a unique one, resulting in 1.2 millions authors (number of papers times average number of authors per paper). The intuition behind the disambiguation process is to reduce this number by merging authors iteratively, based on a list of criteria. That is, for two publications that were thought to belong to two distinct authors, we iteratively merge them if they are likely to have been authored by the same individual. Two authors are considered to be the same individual if all of the following three conditions are fulfilled:

1. Last names of the two authors are identical;
2. Initials of the first names and, when available, given names are the same. If the full first names and given names are present for both authors, they have to be identical;
3. One of the following is true:
    - The two authors cited each other at least once;
    - The two authors share at least one co-author;
    - The two authors share at least one similar affiliations (measured by cosine similarity and tf-idf metrics) (*51*).

The process stops when there is no pair of authors to merge. By applying this disambiguation procedure, we end up with a total of $236,884$ authors, corresponding to a 80% decrease in the number of author names.

Accuracy of the disambiguation algorithm for the APS data

To evaluate the accuracy of our algorithm to disambiguate author names (*47*), we selected 200 pairs of papers that our algorithm predicted to have been written by the same author, and 200 pairs for which the authors are predicted to be distinct individuals. We then determined how many times the pairs correspond to the same individual or not, by searching manually the authors homepage, scholar profile if any, looking at coauthors, affiliation, topic, etc. Out of these 400 pairs, we find the false positive rate (*i.e.* fraction of times the procedure indicates the pair of publications belonging to the same person, while they do not) to be 2% and a false negative rate (*i.e.* fraction of times that the same individual is considered to be two distinct persons) of 12%.

The error rate is likely to be smaller in the subset of authors we explored in this work. Indeed, we consider only authors that have a sufficiently long career, who started around the same period of time and published regularly. We are thus focusing by default on authors for which we have systematic and reliable data, improving, as a consequence, the disambiguation process.

The errors induced by disambiguation are not uniformly distributed among scientists. Indeed, as pointed out in previous research (*52–54*), asian names are the most difficult to disambiguate. Indeed, we examined 500 random pairs of papers from our subset and found that about 90% of manually detected errors are related to Asian names. Therefore, we repeated our analysis by removing from the studied subset all authors with one of the top 60 most common Chinese or Korean last names, as listed in wikipedia (*55, 56*), resulting in 58 careers in total. Removing these examples leads to a further improved disambiguated subset of authors. The analysis performed on this new subset is equivalent to the one reported in the manuscript (see Figs. S2), indicating that our results are robust to the unavoidable small errors induced by disambiguation.

S1.2 Web of Science and Google Scholar data

The APS dataset is highly longitudinal, with complete information about american physicists' publications for over a century. However, it pertains to only one discipline, physics, and only one family of journals, the Physical Review corpus. For this reason we have availed ourselves of Web of Science (WoS) publications and citations to extend our analysis to other disciplines, which we have disambiguated through a number of Google Scholar (GS) profiles data. While automatic name disambiguation in large-scale scholarly datasets remains an unsolved challenge, GS rolled out a scholar profile service, allowing individual scientists to set up, organize and maintain their own publication histories, assisted by Google's algorithmic classification. Hence the GS profiles offer two levels of assurance, making it probably the best disambiguated dataset we can obtain for individual scholars. At the first level, it employs the name disambiguation algorithm developed by the GS team. While the specifics of the algorithms are not available, it is reasonable to assume that Google's algorithms are on par, if not far better than prevailing methods developed by independent academic groups, especially given the open-sourced nature of these academic tools. At the second level, it offers a convenient user interface for individuals to create and maintain their own profiles, hence crowdsourcing the disambiguation effort to ensure the accuracy. This effort also creates a positive feedback loop: as individuals include or exclude publications from their profiles, they effectively help train the model to improve the precision of all other profiles, especially those maintained with less care. Therefore, Google Scholar data offers comprehensive disambiguated individual publication records for a broad range of disciplines.

To this end, we crawled more than 180,000 public profiles from GS. Each profile contains the name of a scientist, a set of keywords identifying his/her research as chosen by the scientist, and the list of his/her publications. Among a scientist's keywords we do not always find the discipline a scientist belongs to. For example, a scientist working in physics might list terms like "neutrinos" or "elementary particles", but not "physics". To identify their discipline, we compile a list of keywords that co-occur in a statistical significant way with the keywords of each discipline. For the six most represented discipline keywords, that is cognitive sciences, chemistry, ecology, economics, biology and neuroscience, we select all profiles that list either

the discipline or one of the statical significant co-occurring keywords. This way we obtained 25,000 profiles in total for the six disciplines.

We then tried to match each publication in each profile with the corresponding one in the WoS database. To do this, we consider a subset of all WoS publications having an author with the same last name as in the GS profile at hand, and calculate the levenshtein distance (*57*) between the title of each GS publication and the publications in the WoS subset. We then consider for the scientist's profile the WoS publication with the lowest levenstein distance (most similar titles) below the threshold value 0.1. If no publication title in the WoS subset has a distance below 0.1, the GS publication is unmatched. We find a WoS match for more than 35% of GS sample. The remainder 65% unmatched publications are due to differences between Google's indexing process and the indexing of Web of Science (Google Scholar lists documents regardless if they are published or not). To check that the high number of unmatched google scholar entries are not due to a wrong matching procedure, we manually checked 200 random unmatched publications, finding that 98% of them are preprints (e.g. arXiv), conference abstracts (e.g. bulletin of the APS), publications in journals or proceedings not indexed by Web of Science, like theses. The remaining 2% unmatched papers are due to changes in the titles that make the algorithm unable to find the right Web of Science database entry.

Finally, we use the citation data from Web of Science, to associate to each matched publication its impact c_{10}. With this methodology we take advantage of both the disambiguation accuracy by Google Scholar and citation indexing accuracy of the Web of Science.

S1.3 Dataset selection

To eliminate authors that leave research at an early stage of their career, in the paper we limit our analysis to scientists that (i) have authored at least one paper every 5 years, (ii) have published at least 10 papers, (iii) their publication career spans at least 20 years in the APS dataset and at least 10 years in the WoS dataset (*58, 59*), arriving to 2,887 scientists in the APS dataset, and 7,630 in the WoS dataset, with persistent publication record. In the following sections, we show that the results of our analysis do not change if different selection criteria or modified filters are used to select the subset of scientists.

Citation-based measures of impact are affected by two major problems: (1) citations follow different dynamics for different papers *(6, 45)* and (2) the average number of citations changes over time (*60*). To overcome (1) for each paper we use the cumulative number of citations the paper received 10 years after its publication, c_{10}, as a measure of its scientific impact *(6, 45, 61)*. We can correct for (2) by normalizing c_{10} by the average c_{10} of papers published in the same year, but this correction does not alter our conclusions (see S1.6), hence we report in the paper results about the APS data without normalization.

To calculate c_{10} we limit the study to publications published up to 2000 for the APS dataset and up to 2002 for the WoS. This requisite together with (iii) implies that the studied scientists started their career in 1980 or before for the APS dataset, or in 1992 or before for the WoS dataset.

S1.4 Choosing different subsets of scientists in the APS dataset

In the main paper, we select APS authors with a long career in science and sustained productivity. More specifically, we restrict our analysis to authors that have *(i)* at least 20 years of career, *(ii)* have at least 10 papers, and *(iii)* have authored at least one paper every 5 years. We expand here our analysis to subsets of APS authors who started their career in different decades and with different career length, showing that our results are valid also for different samples.

Different career length. To ensure that our results are not biased by the specific restrictions on scientific career length, we select three groups of scientists with different career length, that is those that published at least for 5, 10 and 20 years. Note that the restriction on having scientists with at least 5 years career also eliminates most scientists, leaving us with 19,000 individual out of the unique $\sim$230,000 authors of the dataset. Indeed, individuals with a career of less than 5 years represents the 82% of the population, while individuals with more than 20 years of career represent the top 5% of highest career longevity scientists.

Different decades. From the subset of scientists with at least a 10-years long career, we construct four distinct groups: (i) individuals who started their career in between 1950 and 1959 (732 scientists), (ii) in between 1960 and 1969 (1,766 scientists), (iii) in between 1970 and 1979 (2,666 scientists), and finally (iv) in between 1980 and 1989 (3,953 scientists).

Physical Review B dataset Throughout a career, a scientist can publish in different subfields, which might have different typical impact. For this reason, Q might be influenced by a scientist's choice of a sub-area. However, a through study of this hypothesis presents several technical difficulties: (i) detecting fields and understanding their dynamics remains an open research question, since the use of classification scheme like PACS numbers have been shown to be unable to capture emerging 'hot' subfields; (ii) a methodology would be needed to associate an author to just one research sub-field, also accounting for the dynamics of the subfields throughout a career. These technical but crucial aspects, as well as the investigation of all possible scenarios that can arise from a sub-field analysis would require considerable investigation.

However, to probe if subfields dynamics alters our results, we can do a robustness check and consider only papers published in one of the APS journals, namely Physical Review B (PRB), focusing on condensed matter and material physics. The careers reconstructed with this subset of papers are less subjected to subfield changes, since the scope of the PRB is narrower than the entire family of APS journals and can be used as a proxy of subfield. We repeated our key measurements and analysis on the subset of these careers (Fig. S38), and found no qualitative difference with our previous results, indicating that core findings are likely not the artifact of mixing together subfields with different typical impact. However, we do understand that this robustness check does not shed light in the relation between the Q parameter and the change of research subjects throughout a career.

S1.5 Equivalence of APS and WoS data to study physicists

The APS dataset, as provided by the American Physical Society, contains only citations between papers published in the Physical Review journal. That is, if an APS paper is cited by a non-APS paper, this citations will not appear in the dataset. Hence, the number of citations for Physical Review papers is underestimated in the APS dataset. To check whether this could affect our results, we have constructed a dataset where we consider the same scientists and their publication record as described in S1.3, but we quantify paper impact by using citation data from the WoS database. To do this, first we had to find the APS papers in the WoS database through either their DOI, when present or through their title. We succeeded to match publications between APS and WoS in 87% of the cases. We then repeated our key measurements on this new dataset, and found no qualitative differences (Fig. S39). However, the parameters of the Q-model change, since the number of citations c_{10} of papers is systematically larger.

S1.6 Publication impact

Rescaled number of citations. To approximate the scientific impact of each paper, we calculate the number of citations the paper received after 10 years, c_{10}, and we use it as a proxy of publication impact in results reported in the main text. Previous studies *(24, 64, 65)* have shown that the average number of citations per paper changes over time. Indeed for the APS dataset the average number of citations c_{10} fluctuates, while for the WoS dataset it steadily grows (Fig. S3). To be able to compare the impact of papers published at different times and to make sure our results are not affected by this temporal effect, we use a rescaled measure *(24, 60, 66)*, $\tilde{c}^i_{10}$, to gauge the impact of paper i given its publication date d:

$$\tilde{c}^i_{10} = 10 \cdot \frac{c^i_{10}}{\langle c_{10} \rangle_{\delta d}} \tag{S1}$$

where $\langle c_{10} \rangle$ is the average c_{10} calculated over all publications published in the time window $[d - \delta d/2, d + \delta d/2]$. Here we use $\delta d = 1$ year. We find the results of the APS dataset are consistent with the ones based on the raw number of citations c_{10} reported in the main text (Fig. S37). As we find that the rescaling correction does not alter our conclusions, we report in the paper results without normalization, since the raw c_{10} can be better interpreted. For the WoS dataset, we use only the rescaled dataset, since the documented steady growth of the average number of citations can slightly bias the results.

Ranking of papers. The rank of paper in a time window can be used as an alternative proxy of impact. We test that the choice of this measure in the APS dataset to gauge that the specific choice for a paper's impact does not effect our results. We associate each paper with an impact r based on its c_{10}-rank among all the papers published in the same δt window. A paper with $r = 100$ has the highest c_{10} among all papers published in δt, while a paper with, say, $r = 60$ is ranked at the top 40%. We then repeated our measurements and found that our empirical

findings and the resulting random impact rule holds if we convert the impact of each paper c_{10} into rank in a $\delta t = 1$ year window (Fig. S20 and $\delta t = 5$ years window Fig. S21). This result confirms that a scientist's highest impact paper is random in a scientist's career, even when its impact is quantified in terms of rank.

S1.7 Excluding review papers

Review papers follow different statistics *(24, 62, 63)* and could in principle bias our results. For this reason we have repeated our analysis in the APS dataset removing all papers published in Review of Modern Physics, the journal in the APS group containing all review papers (5,123 papers removed). Our results remain unchanged (Fig. S36). Our findings are not altered for two reasons: (i) Review papers are only a small fraction of all papers in the dataset (less than 5%); (ii) Remarkably, they do not change the distribution of highest impact paper c_{10}^*, as shown in Fig. S4. This means that although review papers have a higher average number of citations, they play only a limited role in determining the Q of the individual careers.

S2 Impact patterns preceding and following the highest impact paper

S2.1 Smoothing techniques

The scientists' careers, are represented as time series (Fig. 1A), where each data point corresponds to a publication and the intensity is characterized by its impact. To detect trends before and after the c_{10}^* peak, we apply two standard techniques to these time series (*67*). The first technique is moving average, which provides a series of averages over different windows of the original time series. Given a series of numbers and a fixed window of size L, the first element of the moving average is obtained by taking the average of the initial L numbers in the series. Then the window is modified by "shifting it forward", that is, excluding the first number of the series and including the next number following the original subset in the series. This process is repeated over the entire time series, finally providing a new time series made of all averages, where short-term fluctuations are smoothed out. The second technique used is a record series. Similar to the moving average, it produces a new time series by rolling a window L on the original data. The difference between this method and moving average is that the maximum value of the L numbers is considered in this case.

We apply both techniques using different values of L ($L = 1, 2, 5, 10, 20$) to all points but c^* of individual time series. For each resulting set of time series, we consider the average $\langle c_{10} \rangle$ before and after t^*, time of the highest impact paper, for individuals having similar c_{10}^* (Fig. S6). We consider three groups of scientists, segmenting them into low, middle and high impact groups based on c_{10}^*. For each group, we first compute the average c_{10}^*. We then compute the average value of the points immediately preceding c_{10}^*, the average value of the second points

preceding c_{10}^*, etc., as well as the average value of the first points following c_{10}^*, the average value of the second points following c_{10}^*, and so on. We find the results in Fig. 2B are robust against the choice of L, and we chose $L = 10$ to compute Fig. 2B.

S2.2 Detection of trends preceding and following the highest impact paper

The observed plateaus before and after the highest impact paper in Fig. S6 can be the result of different trends in individual careers that cancel out when computing the average. To test this hypothesis we analyze the individual trend before and after c_{10}^*. For each scientist i, we fit the paper impact c_{10} over time before t^* with the function:

$$c_{10}^i(t)_{t<t^*} \sim \beta_i t \tag{S2}$$

Similarly, we fit the trend after t^* with the function:

$$c_{10}^i(t)_{t>t^*} \sim \alpha_i t \tag{S3}$$

An overall increase in the impact of papers before the highest impact paper corresponds to a positive β, while a negative α indicates a decay in the impact after the highest impact paper. The fit is performed by using a least square method on the time series smoothed with a moving average with $L = 10$; The resulting distributions of β and α ($P(\beta)$ and $P(\alpha)$), illustrate the variabilities in slopes across the entire data set (Fig S7A-B). The average of $P(\beta)$ is $\langle\beta\rangle = 0.11$ and the distribution is approximately symmetric around the peak. The average of $P(\alpha)$, $\langle\alpha\rangle = 0.07$ and the distribution is shifted towards the left, indicating that for a large fraction of scientists α is negative. However, to see whether these trends are induced by the highest impact paper, we repeat the measurements of α and β on randomized careers, where we shuffle the impact of every paper within a scientist's career but preserve the time of each publication and c_{10}^*, the impact of the highest impact paper. We find that the distributions of trends preceding and following the highest impact paper are similar, as confirmed by Mann-Whitney U tests (with p-value $p = 0.15$ and $p = 0.34$ for $P(\beta)$ and $P(\alpha)$ respectively). We also investigate whether researchers with different c_{10}^* exhibit different trends. To this end, we measure $\langle\alpha\rangle$ and $\langle\beta\rangle$ as a function of c_{10}^* (Fig. S7C-D), and find that there are not significant variations between scientist with different c_{10}^*. Taken together, Fig. S7 indicates that the observed trends preceding and following the highest impact paper are not different from those of randomized careers, further supporting our conclusion that random is impact within a scientific career, even before and after the highest impact paper.

S3 Random Impact rule

S3.1 Career patterns and the random impact rule in the literature

The association between a person's age and exceptional accomplishment is a question that has for centuries fascinated many researchers, dating back to at least 1874 when Beard presented the relation of work to age and estimated that peak performance in science and creative arts typically occurred between the ages of 35 and 40 (*68*). Throughout all relevant studies, no matter what domain we focus on and how we define achievement, the probability to succeed is a curvilinear, single-peak function of age *(21, 69–73)*, documenting the existence of a peak age of a career. While the precise location of peak age can shift over time *(16, 74)*,differ across individuals (*75*) and change across disciplines *(20, 72, 76, 77)*, one remarkably robust pattern is that there is always a peak age and it always occurs around the 30s to 40s in a lifecycle. Indeed, the lifecycle of a career is characterized by the rarity of contributions at the beginning of life and a gradual decline in mid to late life cycle. The absence of output in early age is primarily driven by schooling and training (*73, 74*). That is, if one wants to stand on the shoulders of giants, one must first climb their backs innovating in a meaningful manner. The decline in the frequency of great scientific breakthroughs in mid to late life cycle can be attributed to many factors, such as the obsolescence of skill, degradation in health that limits productivity, increasing preferences towards retirement, family responsibilities, demanding administrative duties, and more. Yet, surprisingly or not, although very plausible, none of these studies have tested changes in productivity as an underlying explanation of when outstanding achievements occurs in a scientific career.

S3.2 Random impact rule in the WoS dataset

Our findings about the random impact rule, described in the main paper, are also generalizable to disciplines other than physics. The results of our empirical analysis for the six additional disciplines of the WoS dataset are reported in Figs. S23-S22 We find that the random impact rule is universal, being robust across all the studied disciplines.

S3.3 Random impact rule for different subsets of APS scientists

We have studied the random impact rule in the datasets described in S1.4. Our empirical findings hold across different career lengths (Figs. S13, S14) and for scientists that start their career in different decades (Figs. S15, S16) indicating that our restrictions to the subset do not bias our results. Note that the specific shape of the timing of the highest impact paper, $P\left(t^*\right)$, changes over different career length or decades. However, what does not change is the lack of differences between the measurement in the original and in the randomized data, which confirms that the random impact rule is a characteristic of scientific careers, regardless of their length or of their starting time.

S3.4 Impact autocorrelation

To provide further evidence that impact is random within a scientific career, we measure the autocorrelation function of the impact of a scientist's sequence of papers. Given a sequence of N papers for a scientist, we measure:

$$\rho(\tau) = \frac{\langle c_{10}(n)c_{10}(n+\tau)\rangle - \langle c_{10}\rangle^2}{\sigma^2} \tag{S4}$$

where n is the position of the paper in the sequence, τ is the position difference of two papers in the sequence, σ^2 is the variance of the distribution of c_{10}, and the average $\langle\ \rangle$ is performed by rolling a window of size τ over the scientist's sequence of publication. For example, two consecutive papers have $\tau = 1$, and there are $N-1$ pairs of papers with $\tau = 1$ in a sequence of N publications, while the first and the last paper in a sequence are the only pair of papers for which $\tau = N-1$. In a random process, the autocorrelation function is zero for any $\tau \neq 0$. We report the autocorrelation function ρ as a function of the lag τ in Fig. S25 and Fig. S26 for the APS dataset and WoS dataset, respectively. The extremely small values of autocorrelation ($\|\rho(\tau)\| < 0.05$, $\forall\tau$) indicate that close papers in a scientist's sequence of publication do not have similar impact, further supporting the random impact rule.

S4 Models

S4.1 Fitting the impact distribution $P\left(c_{10}\right)$

Two family of models have been widely proposed for citation distributions: power-law (*61, 78*) and lognormal *(24, 62, 82)*. If the variable c_{10} follows a lognormal distribution, then $P\left(c_{10}\right)$ is written as

$$P\left(c_{10}\right) = \frac{1}{c_{10}\sqrt{2\pi}\sigma} e^{-\frac{(\log c_{10}-\mu)^2}{2\sigma^2}} \tag{S5}$$

where μ and σ are the two parameters of the distribution. If instead the variable c_{10} is distributed according to a power-law, then

$$P\left(c_{10}\right) = \frac{\gamma-1}{\left(c_{10,\text{min}}\right)^{1-\gamma}} c_{10}^{-\gamma} \tag{S6}$$

where γ is the parameter of the distribution and $c_{10,\text{min}}$ is the lower bound for the power-law behaviour. Here we want to compare these two models to see which one is the better fit for the empirical distribution $P\left(c_{10}\right)$ found in our data. To do this, we first estimate with a maximum likelihood approach the parameters μ and σ of the lognormal model (*80*), and γ and c_{10}^{min} of the power-law model (*81*) that best fit our data. We then compare the two by calculating their p-value using a Kolmogorov-Smirnov statistics approach, as described in (*81*). We find that $\mu = 1.93$ and $\sigma = 1.05$, while $\gamma = 3.13$ and $c_{10}^{\text{min}} = 49$. Using these parameters, we can

generate two different kinds of synthetic data and see which one represents a better fit of the empirical data. By doing this, we can rule out the power-law model in favour of the lognormal model (see also Fig. S27).

S4.2 R-model and its predictions

The Random Impact Model (R-model) assumes that each scientist publishes a sequence of N papers whose impact is randomly chosen from the same impact distribution $P(c_{10})$. We do not assume any model for the productivity, rather we use the observed productivity distribution $P(N)$. For any non-singular function $P(c_{10})$, a productive scientist has a higher probability to score a high c_{10}^*, as it has more chances to draw a large c_{10}. To quantify this, we calculate the expectation value of the largest c_{10}, $\langle c_{10}^* \rangle$, as a function of the number of publications N. The value c_{10}^* is the largest one if there are no others greater than it. The probability for there to be no value larger than c_{10}^* is given by the cumulative distribution function $\int_0^{c_{10}^*} P(c'_{10})\mathrm{d}c'_{10}$. Since a scientist has N publications in total, all the other $(N-1)$ need to be smaller than c_{10}^*, the probability of this event being $\left(\int_0^{c_{10}^*} P(c'_{10})\mathrm{d}c'_{10}\right)^{N-1}$. Since there are N possibilities to extract c_{10}^*, each with probability $P(c_{10}^*)$, the overall probability $\pi(c_{10}^*)$ that c_{10}^* is the largest value is

$$\pi(c_{10}^*) = NP(c_{10}^*)\left[\int_0^{c_{10}^*} P(c'_{10})\mathrm{d}c'_{10}\right]^{(N-1)} \tag{S7}$$

Finally, the expected value of c_{10}^* is then:

$$\langle c_{10}^* \rangle (N) = \int_0^{\infty} c_{10}^* \pi(c_{10}^*)\mathrm{d}c_{10}^*. \tag{S8}$$

Identical arguments hold to calculate $\langle \log c_{10}^* \rangle (N)$, where it is sufficient to substitute $P(c_{10})$ with $P(\log c_{10})$. In our case, $P(c_{10})$ is a lognormal function (see S4.1 and Fig. 2A), or equivalently, $P(\log c_{10})$ is a normal distribution. Unfortunately, no explicit form is known to express the cumulative distribution $\int_{c_{10}^*}^{\infty} P(c'_{10})\mathrm{d}c'_{10}$ and, consequently, the expected largest value $\langle c_{10}^* \rangle$[1] (*82*). However, (S8) can be easily solved numerically, and it is reported in Fig. 2C. We find that R-model prediction offers a poor description of the data.

We are also interested to know how the magnitude of the highest impact publication c_{10}^* scales with the average impact calculated over all other publications, $\langle c_{10}^{-*} \rangle$. The average $\langle c_{10}^{-*} \rangle$ of N numbers extracted from a distribution $P(c_{10})$ is defined as

$$\langle c_{10}^{-*} \rangle = \frac{\int_0^{c_{10}^*} c_{10} P(c_{10})\,\mathrm{d}c_{10}}{\int_0^{c_{10}^*} P(c_{10})\,\mathrm{d}c_{10}} \tag{S9}$$

[1] An approximate solution can be obtained for large N by using the Fisher-Tippet-Gnedenko theorem. However, since in our dataset many scientists have productivity $N < 100$, this asymptotic solution is inadequate to capture the behavior of $c_{10}^*(N)$.

As before, the scaling of $\langle \log c_{10}^{-*} \rangle$ with $\log c_{10}^{*}$ can be calculate by substituting $P(c_{10})$ with $P(\log c_{10})$. Eq. (S9) can be solved numerically, and gives a monotonic relation between $\langle c_{10}^{-*} \rangle$ and $\langle c_{10}^{*} \rangle$. Because of the log-normal nature of $P(c_{10})$, for high N the first moment $\langle c_{10} \rangle$ of the distribution converges, while its maximum expected value, which determines c_{10}^{*}, continues to increase with N. This behavior is evident in Fig. 3D. Hence beyond $\langle c_{10}^{-*} \rangle \sim 1.97$ the R-model cannot discriminate between high impact scientists.

S4.3 The Q-model

A lognormal distribution is often a signature of a multiplicative process (*83, 84*). For this reason, we write the impact c_{10}, which is log-normal distributed (Fig. 2A), as the product of two terms, the Q parameter and paper potential impact. More specifically, the impact $c_{10,i\alpha}$ of paper α published by scientist i is

$$c_{10,i\alpha} = Q_i p_\alpha. \tag{S10}$$

where p_α encodes the potential impact of paper α, while Q_i is an individual parameter for scientist i. We assume that p_α is drawn from a distribution $P(p)$ that, as shown below, turns to be the same for all scientists and captures comparable access to ideas and resources. The parameter Q_i captures the ability of scientist i to take advantage of the available information, enhancing (or diminishing) the impact of paper α. Q_i is considered to be constant throughout a scientist's career. The resulting Q-model assumes that each scientist selects a project p_α from a distribution $P(p)$ that is the same for all scientists, and improves on it with a factor Q_i that is unique to the scientist, resulting in a paper of impact (S10).

The stochastic process behind the Q-model is determined by the joint probability $P(p, Q, N)$, which in general contains all pairwise correlations between the variables p, Q and N. The log-normal nature of $P(c_{10})$ (Fig. 2A) allows us to measure $P(p)$. Indeed, from (S10) for each paper α of scientist i we have $\log p_\alpha = \log c_{10,i\alpha} - \log Q_i$. Since we make the hypothesis that Q_i is constant throughout the entire career of a scientist (which we confirm later), Q can be estimated by averaging over all publications of the scientist, obtaining $\log Q_i = \langle \log c_{i,10} \rangle - \mu_p$, where we assumed the average potential impact of the i' publications to be equal to the mean of $P(\hat{p})$, $\langle \hat{p} \rangle = \mu_p$. We can then look at the distribution of $\log c_{10,i\alpha} - \langle \log c_{i,10} \rangle$ across all publications of all scientists as a proxy of the distribution of $\log p$ (see Fig. 2F). The distribution has average zero, as in the expression of Q_i we disregard the common translational factor μ_p, which we cannot estimate at this point. Nevertheless, the measurement of this distribution allows us to assume correctly that the distribution of $P(p)$ is lognormal. Since we can measure directly the productivity distribution $P(N)$ and find that it follows also a lognormal, we assume that $P(p, Q, N) = \log \mathcal{N}(\mu, \Sigma)$ is overall a trivariate lognormal distribution with mean μ and co-variance matrix Σ.

S4.4 Maximum-Likelihood estimation of the Q-model parameters

We estimate the parameters μ and Σ of the trivariate lognormal distribution $P(p, Q, N)$ from the data by using a maximum-likelihood approach. A trivariate lognormal distribution is a trivariate normal distribution of the logarithms of its variables (*85*). Since in the usual parameterization of a lognormal μ and Σ indicate the average vector and the co-variance matrix of the logarithm of the variables, estimating these parameters is equivalent to the parameters estimation for the corresponding trivariate normal distribution. For convenience, we work with the logarithm of the variables, denoted with $\hat{p} = \log p$, $\hat{Q} = \log Q$ and $\hat{N} = \log N$. We have then

$$P\left(\hat{p}, \hat{Q}, \hat{N}\right) = \frac{1}{\sqrt{(2\pi)^3}} \exp -\frac{1}{2} \left(\mathbf{X} - \mu\right)^T \Sigma^{-1} \left(\mathbf{X} - \mu\right). \tag{S11}$$

where $\mathbf{X}$ and μ are vectors of three elements, namely $\mathbf{X} = \left(\hat{p}, \hat{Q}, \hat{N}\right)$ and $\mu = (\mu_p, \mu_Q, \mu_N)$, while Σ is a 3×3 matrix,

$$\Sigma \equiv \begin{pmatrix} \sigma_p^2 & \sigma_{p,Q} & \sigma_{p,N} \\ \sigma_{p,Q} & \sigma_Q^2 & \sigma_{Q,N} \\ \sigma_{p,N} & \sigma_{Q,N} & \sigma_N^2 \end{pmatrix}, \tag{S12}$$

which contains the variances σ_p^2, σ_Q^2 and σ_N^2 of the three marginal distributions $P(p)$, $P(Q)$, $P(N)$ respectively, and all the mixed terms $\sigma_{p,Q}$, $\sigma_{p,N}$ and $\sigma_{Q,N}$. Σ is a symmetric positive definite matrix.

Given a scientist i, the joint probability that s/he has a productivity $\hat{N}_i$, individual parameter $\hat{Q}_i$ and a paper α with impact $c_{i\alpha}$, we can rewrite the joint probability function (S11) without the dependence on $\hat{p}$ as

$$\begin{aligned} P \quad & (\log c_{i\alpha} - \hat{Q}_i, \hat{Q}_i, \hat{N}_i) = \\ = \quad & \frac{1}{\sqrt{(2\pi)^3 \left|\Sigma\right|}} \exp \left\{ -\frac{1}{2} \left[\begin{pmatrix} \log c_{i\alpha} - \hat{Q}_i \\ Q_i \\ \hat{N} \end{pmatrix} - \begin{pmatrix} \mu_p \\ \mu_Q \\ \mu_N \end{pmatrix} \right]^T \Sigma^{-1} \left[\begin{pmatrix} \log c_{i\alpha} - \hat{Q}_i \\ \hat{Q}_i \\ \hat{N} \end{pmatrix} - \begin{pmatrix} \mu_p \\ \mu_Q \\ \mu_N \end{pmatrix} \right] \right\}, \end{aligned} \tag{S13}$$

where we used the relation $\hat{p_\alpha} = \log p_\alpha = \log c_{10,i\alpha} - \log Q_i = \log c_{10,i\alpha} - \hat{Q}_i$, simply derived from Eq. (S10). The inverse of the covariance matrix, Σ^{-1} reads

$$\Sigma^{-1} = \frac{1}{K} \begin{pmatrix} \sigma_Q^2\sigma_N^2 - \sigma_{Q,N}^2 & -\sigma_{p,Q}\sigma_N^2 + \sigma_{p,N}\sigma_{Q,N} & \sigma_{p,Q}\sigma_{Q,N} - \sigma_{p,N}\sigma_Q^2 \\ -\sigma_{p,Q}\sigma_N^2 + \sigma_{p,N}\sigma_{Q,N} & \sigma_p^2\sigma_N^2 - \sigma_{p,N}^2 & -\sigma_p^2\sigma_{Q,N} + \sigma_{p,N}\sigma_{p,Q} \\ \sigma_{p,Q}\sigma_{Q,N} - \sigma_{p,N}\sigma_Q^2 & -\sigma_p^2\sigma_{Q,N} + \sigma_{p,N}\sigma_{p,Q} & \sigma_Q^2\sigma_N^2 - \sigma_{Q,N}^2 \end{pmatrix} \tag{S14}$$

with K being that determinant of Σ

$$K = \sigma_p^2\sigma_Q^2\sigma_N^2 + 2\sigma_{p,Q}\sigma_{Q,N}\sigma_{p,N} - \sigma_{p,N}^2\sigma_Q^2 - \sigma_{Q,N}^2\sigma_p^2 - \sigma_{p,Q}^2\sigma_N^2.$$

To obtain the marginal distribution over a subset of multivariate normal random variables, one only needs to drop the irrelevant variables (the variables that one wants to marginalize out) from the mean vector and the covariance matrix (*85*). It is therefore straightforward to derive from Eq. (S13) the marginal $P(Q,N)$, expressing the probability to have a scientist with productivity N and parameter Q. We can then write the probability that s/he a sequence of publications with potential impact $\{\log c_{i\alpha}\} - \hat{Q}_i$ as

$$\begin{aligned}\mathcal{L}_i &= \int P\left(\hat{Q}_i,\hat{N}_i\right)\prod_{\alpha=1}^{\exp(\hat{N}_i)} P\left(\log c_{i\alpha}-\hat{Q}_i|\hat{Q}_i,\hat{N}_i\right)\mathrm{d}\hat{Q}_i = \\ &= \int P\left(\hat{Q}_i,\hat{N}_i\right)^{(1-\exp\hat{N}_i)}\prod_{\alpha=1}^{\exp(\hat{N}_i)} P\left(\log c_{i\alpha}-\hat{Q}_i,\hat{Q}_i,\hat{N}_i\right)\mathrm{d}\hat{Q}_i \qquad \text{(S15)}\end{aligned}$$

where $P\left(\log c_{i\alpha}-\hat{Q}_i|\hat{Q}_i,\hat{N}_i\right)$ is the conditional probability that scientist i has a publication with potential impact $\log c_{i\alpha}-\hat{Q}_i$ given his/her N_i and Q_i, and we have $P\left(\log c_{i\alpha}-\hat{Q}_i|\hat{Q}_i,\hat{N}_i\right) = \frac{P\left(\log c_{i\alpha}-\hat{Q}_i,Q_i,N_i\right)}{P(Q_i,N_i)}$. In Eq. (S15) we have also used the result about the lack of impact correlations between a scientist's publications to write the probability of the publications sequence as a product of independent events $\prod_{\alpha=1}^{\exp(\hat{N}_i)} P\left(\log c_{i\alpha}-\hat{Q}_i|\hat{Q}_i,\hat{N}_i\right)$. Finally to get the likelihood, one has to integrate over all possible Q_i. Using Eqs. (S13) and (S14) in (S15) we obtain

$$\begin{aligned}\mathcal{L}_i &= \int \frac{1}{D_i} e^{-\left(A_i\hat{Q}_i{}^2+B_i\hat{Q}_i+C_i\right)}\mathrm{d}\hat{Q}_i = \\ &= \frac{1}{D_i}\sqrt{\frac{\pi}{A_i}}\, e^{\frac{B_i^2}{4A_i}-C_i} \qquad \text{(S16)}\end{aligned}$$

where

$$\begin{aligned}A_i &= \frac{1}{2}\left\{\frac{1-\exp\hat{N}_i}{\sigma_Q^2\sigma_N^2-\sigma_{Q,N}^2}\sigma_N^2+\frac{\exp\hat{N}_i}{K}\left[\sigma_Q^2\sigma_N^2-\sigma_{Q,N}^2-2\left(\sigma_{p,N}\sigma_{Q,N}-\sigma_{p,Q}\sigma_N^2\right)+\sigma_p^2\sigma_N^2-\sigma_{p,N}^2\right]\right\} \qquad \text{(S}\\ B_i &= \frac{1}{2}\left\{\frac{1-\exp\hat{N}_i}{\sigma_Q^2\sigma_N^2-\sigma_{Q,N}^2}\left[-2\mu_Q\sigma_N^2-2\sigma_{Q,N}(\hat{N}_i-\mu_N)\right]+\frac{\exp\hat{N}_i}{K}\left[-2\left(\frac{\sum_\alpha\log c_{i\alpha}}{\exp\hat{N}_i}-\mu_p\right)\left(\sigma_Q^2\sigma_N^2-\sigma_{Q,N}^2\right)\right.\right.\\ &+ 2\left(\frac{\sum_\alpha\log c_{i\alpha}}{\exp\hat{N}_i}-\mu_p+\mu_Q\right)\left(\sigma_{p,N}\sigma_{Q,N}-\sigma_{p,Q}\sigma_N^2\right)-2\left(\sigma_{p,Q}\sigma_{Q,N}-\sigma_{p,N}\sigma_Q^2\right)\left(\hat{N}_i-\mu_N\right)+\\ &- \left.\left.2\mu_Q\left(\sigma_p^2\sigma_N^2-\sigma_{p,N}^2\right)+2\left(\sigma_{p,N}\sigma_{p,Q}-\sigma_p^2\sigma_{Q,N}\right)\left(\hat{N}_i-\mu_N\right)\right]\right\} \qquad \text{(S}\end{aligned}$$

$$
\begin{aligned}
C_i &= \frac{1}{2}\left\{\frac{1-\exp\hat{N}_i}{\sigma_Q^2\sigma_N^2-\sigma_{Q,N}^2}\left[\sigma_N^2\mu_Q^2+2\sigma_{Q,N}\mu_Q\left(\hat{N}_i-\mu_N\right)+\sigma_Q^2\left(\hat{N}_i-\mu_N\right)^2\right]+\right.\\
&+ \frac{\exp\hat{N}_i}{K}\left[\left(\sigma_Q^2\sigma_N^2-\sigma_{Q,N}^2\right)\left(\frac{\sum_\alpha(\log c_{i\alpha})^2}{\exp\hat{N}_i}+\mu_p^2-2\mu_p\sum_\alpha\log c_{i\alpha}\right)-\right.\\
&+ 2\mu_Q\left(\sigma_{p,N}\sigma_{Q,N}-\sigma_{p,Q}\sigma_N^2\right)\left(\frac{\sum_\alpha\log c_{i\alpha}}{\exp\hat{N}_i}-\mu_p\right)+\\
&+ 2\left(\sigma_{p,Q}\sigma_{Q,N}-\sigma_{p,N}\sigma_Q^2\right)\left(\hat{N}_i-\mu_N\right)\left(\frac{\sum_\alpha\log c_{i\alpha}}{\exp\hat{N}_i}-\mu_p\right)+\mu_Q^2\left(\sigma_p^2\sigma_N^2-\sigma_{p,N}^2\right)+\\
&- \left.\left.2\mu_Q\left(\sigma_{p,N}\sigma_{p,Q}-\sigma_p^2\sigma_{Q,N}\right)\left(\hat{N}_i-\mu_N\right)+\left(\sigma_p^2\sigma_Q^2-\sigma_{p,Q}^2\right)\left(\hat{N}_i-\mu_N\right)^2\right]\right\}\\
D_i &= (2\pi)^{\left(1+\frac{1}{2}\exp\hat{N}_i\right)}\left(\sigma_Q^2\sigma_N^2-\sigma_{Q,N}^2\right)^{\frac{1-\exp\hat{N}_i}{2}}|K|^{\frac{\exp\hat{N}_i}{2}}
\end{aligned}
$$

Eventually, the total log-likelihood $\log\mathcal{L}$ for the set of the M scientists of the dataset is

$$
\begin{aligned}
\log\mathcal{L} &= \sum_{i=1}^{M}\log\mathcal{L}_i = \\
&= \sum_{i=1}^{M}\left[\left(\frac{B_i^2}{4A_i}-C_i\right)+\log\frac{1}{D_i}\sqrt{\frac{\pi}{A_i}}\right]
\end{aligned}
\tag{S19}
$$

$\log\mathcal{L}$ is just a function of the parameters μ and Σ, which can be easily estimated by calculating the log-likelihood (S19) based on the impact of the 106,463 publications of the $2,887$ scientists of the dataset and finding the values of μ and Σ for which $\log\mathcal{L}(\mu,\Sigma)$ has a maximum. The search of the maximum has been performed by using the fmincon function in the Optimization toolbox of Matlab, providing it 10 different initial conditions and running the optimization function 10 times for each of them. As final estimation of the parameters, we take the average of all the outputs, obtaining:

$$
\mu=(0.92,0.93,3.34)\,,\ \Sigma=\begin{pmatrix}0.93 & 10^{-3} & 7\cdot10^{-3}\\ 10^{-3} & 0.21 & 0.09\\ 7\cdot10^{-3} & 0.09 & 0.33\end{pmatrix}. \tag{S20}
$$

The estimation of the parameters μ_N and σ_N^2 is consistent with the direct measurement done fitting the marginal $P(N)$ (Fig. 2B), which is a lognormal with $\mu=3.6$ and $\sigma^2=0.57$. The estimation of μ_Q, μ_p, σ_Q^2 and σ_p^2 also agrees closely with the statistical moments of $P(c_{10})$

($\mu = 1.93$ and σ^2=1.05, see Fig. 2A): given Eq. (S10) and that the sum of two random normal variables follows a normal distribution with mean equal to the sum of the two means and variance equal to the sum of the two variances, we have $\mu_{c_{10}} = \mu_p + \mu_Q = 1.85$ and $\sigma^2_{c_{10}} = \sigma^2_p + \sigma^2_Q = 1.14$. However, the most important prediction of (S20) is given by the estimation of the mixed terms $\sigma_{p,N}$ and $\sigma_{p,Q}$ which are respectively one and two orders of magnitude smaller than the other mixed terms and the diagonal terms respectively. Indeed, this prediction implies that the joint probability (S11) factorizes as $P\left(\hat{p}, \hat{Q}, \hat{N}\right) \simeq P\left(\hat{p}\right) P\left(\hat{Q}, \hat{N}\right)$, meaning that the probability to extract a project of potential impact $\hat{p}$ is indeed the same for all scientists, regardless of the individual parameter Q or productivity. The positive value of the term $\sigma_{Q,N}$ indicates instead that high Q is slightly associated with high productivity.

S4.5 Computing a scientist's Q parameter

The approach described in the previous section allows us to compute the individual Q_i, which corresponds to the value maximizing the individual likelihood Eq. (S16). To find this maximum, we rewrite the likelihood $\mathcal{L}_i$ in a simpler form by using the factorization $P\left(\hat{p}, \hat{Q}, \hat{N}\right) = P\left(\hat{p}\right) P\left(\hat{Q}, \hat{N}\right)$,

$$\mathcal{L}_i(\hat{Q}_i) = P(\hat{Q}_i, \hat{N}_i) \prod_{\alpha=1}^{N_i} P(\log c_{i\alpha} - \hat{Q}_i). \tag{S21}$$

The log-likelihood function follows

$$\log \mathcal{L}_i(\hat{Q}_i) = \log P(\hat{Q}_i, \hat{N}_i) + \sum_{\alpha=1}^{N_i} \log P(\log c_{i\alpha} - \hat{Q}_i), \tag{S22}$$

where we have

$$\log P(\log c_{i\alpha} - \hat{Q}_i) = -\frac{1}{2\sigma_p^2}(\hat{p} - \mu_p)^2 - \log \sqrt{2\pi\sigma_p^2}, \tag{S23}$$

and

$$\log P(\hat{Q}, \hat{N}) = -\frac{1}{2\left|\Sigma_{QN}\right|}\left[\sigma_N^2(\hat{Q} - \mu_Q)^2 - 2\sigma_{QN}\hat{Q}_i(\hat{N}_i - \mu_N)\right] - \log\left(2\pi\left|\Sigma_{QN}\right|\right) \tag{S24}$$

with

$$|\Sigma_{QN}| = \sigma_Q^2\sigma_N^2 - \sigma_{QN}^2. \tag{S25}$$

Using (S23) and (S24) in (S22), we obtain

$$-\log \mathcal{L}_i = \frac{1}{2\sigma_p^2}\sum_{\alpha=1}^{N_i}(\log c_{i\alpha} - \hat{Q}_i - \mu_p)^2 + \frac{1}{2\left|\Sigma_{QN}\right|}\left[\sigma_N^2(\hat{Q}_i - \mu_Q)^2 - 2\sigma_{QN}\hat{Q}_i(\hat{N}_i - \mu_N)\right] + \text{const},$$

where const indicates a term not depending on Q_i. We finally find that the log-likelihood can be written in the form $\log \mathcal{L}_i = -A_i \hat{Q}_i^2 - B_i \hat{Q}_i + \text{const}$, with

$$A_i = \frac{N_i}{2\sigma_p^2} + \frac{\sigma_N^2}{2\left|\Sigma_{QN}\right|} \tag{S26}$$

and

$$B_i = -\frac{N_i\left(\langle \log c_{i\alpha} \rangle - \mu_p\right)}{\sigma_p^2} - \frac{\sigma_N^2 \mu_Q + \sigma_{QN}(\hat{N}_i - \mu_N)}{\left|\Sigma_{QN}\right|} \tag{S27}$$

Finally, the maximum of $\log \mathcal{L}_i$ is given by

$$\begin{aligned} \hat{Q}_i &= -\frac{B_i}{2A_i} = \\ &= \frac{\langle \log c_{i\alpha} \rangle - \mu_p + \frac{\sigma_N^2 \sigma_p^2 \mu_Q + \sigma_{QN}\sigma_p^2(\hat{N}_i - \mu_N)}{N_i\left|\Sigma_{QN}\right|}}{1 + \frac{\sigma_N^2 \sigma_p^2}{N_i\left|\Sigma_{QN}\right|}} \end{aligned} \tag{S28}$$

which expanding using the relation $\frac{1}{1+x} = 1 - x + \mathcal{O}\left(x^2\right)$ can be approximated as

$$\hat{Q}_i = \langle \log c_{i\alpha} \rangle - \mu_p + \mathcal{O}\left(\frac{1}{N_i}\right). \tag{S29}$$

Hence, for sufficiently large N_i, we have

$$Q_i = e^{\langle \log c_{i\alpha} \rangle - \mu_p}, \tag{S30}$$

which is the Q estimation reported in Eq. (4) of the main text. This estimation correlates very well with the exact form of Q, calculated based on (S28) (Fig. S28).

S4.6 Goodness of the Q-model

We have shown that the Q-model represents a much better fit to the data than the R-model (Figs. 2C-D). Here we test the goodness of fit of the Q-model to the data. Since the Q-model relies on the estimation of the parameters of a multivariate distribution, it is not appropriate to use approaches based on the "distance" between the observed distribution and the estimated distribution, like the KS-statistics described in (*81*) and used in S4.1. Therefore, we assess the goodness of the Q-model to fit the data by using its predictions. More specifically, based on the Q-model we can calculate analytically how $\langle \log c_{10}^* \rangle$ varies as a function of the productivity N (Eq. (S35)) and as a function of the average impact without the highest impact paper $\langle c_{10}^{-*} \rangle$ (Eq. (S36)). Note that $\langle c_{10}^{-*} \rangle$ is a proxy of the Q-parameter when N is large enough (Eq. (S30)). To evaluate the goodness of these predictions, we calculate the coefficient of determination R^2

between the prediction of the Q-model (red line in Fig. S29A-B) and the same quantity estimated from the data (grey circles in Fig. S29A-B). We find that $R^2 = 0.97$ for $\langle \log c_{10}^* \rangle \, (N)$ and $R^2 = 0.99$ for $\langle \log c_{10}^* \rangle \left(\langle c_{10}^{-*} \rangle \right)$. We also compute the coefficient of determination R^2_{scatt} between the Q-model prediction and the scattered data (grey stars in the background of Fig. S29A-B). Due to the stochastic nature of the model, there is an inherent deviation within the Q-model prediction between the scattered data points and the mean. For example when we fix N, the variability of c_{10}^* is, to a large extent, due to scientists having different Q. Similarly, when we fix $\langle c_{10}^{-*} \rangle$, which is equivalent to fixing Q when N is large, the variability in c_{10}^* is due to the different productivity N of scientists. To account for such stochasticity, we generate synthetic data by means of the joint probability Eq. (S11) with the estimated parameters.

More specifically, we generate 1000 instances of synthetic data, each instance being made of 2887 careers, as in the original data, each career having a productivity N_i and parameter Q_i extracted from $P\left(Q,N\right)$ and a sequence of impact $c_{10,i\alpha}$ obtained by extracting randomly p_α from $P(p)$ and multiplying it by Q_i. We then compute for each of these synthetic datasets the R^2_{synth} (coefficient of discrimination between the scattered data and the Q-model prediction) and find that the R^2_{scatt} is contained in the distribution $P\left(R^2_{\text{synth}}\right)$ (Fig. S29C-D), which indicates that the synthetic data generated through the model are scattered in the same way as the original data. In addition to the R^2 we also compare the distribution of residuals for original and synthetic data, and find that the distributions are indistinguishable (Fig. S29E-F). Taken together, Fig. S29 provides quantitative evidence on the goodness of fit of the Q-model to the data.

Finally, a fundamental prediction of the Q-model is that the rescaled variables $c_{10,i\alpha}/Q_i$ collapse on the universal distribution $P(p)$. The collapse of different probability distributions onto a single curve is a standard concept in statistical physics, representing the existence of a universal law underlying a phenomenon. We demonstrate the evidence of this data collapse in Fig. 3C of the main text, documenting the validity of the Q-model.

S4.7 Q-model and the real data-generating process

The Q-model treats each observation as a triple of (p, Q, N) drawn from $P(p, Q, N)$. The real data-generating process assumes that we draw a scientist with productivity N and Q, and then that scientist makes N draws from the potential impact distribution $P(p)$, which could depend (in principle) on Q and N. The Q-model and this latter process are equivalent thanks to the documented uncoupling between p and (Q, N) (see Eq. (2) of the main text).

S4.8 Q-model for the WoS dataset

In the main paper we show results and predictions of the Q-model for the APS dataset. We have also tested the hypothesis and predictions of the Q-model on the newly curated datasets, finding again excellent agreement between model and data. The only difference across disciplines is that we obtained slightly different parameters for different fields, which is expected (see Fig. S30).

S4.9 Stability and measurement accuracy of the Q-parameter

Given the stochastic nature of the Q-parameter, we are prompted to scrutinize the stability of the parameter over time and with which accuracy we can measure it early in a career. Indeed, the Q-parameter might change over time: in Fig. S31A we study the stability of the Q-parameter by applying Eq. (S28) to a moving window of ΔN papers and measuring $Q\left(\Delta N\right)$. The observed changes in the estimated value are due to two reasons:

1. stochastic uncertainty, ΔQ, due to the fact that we estimate the value of Q by using a finite number of points (accuracy).

2. the Q-parameter is not a perfect parametrization of the scientist's career and varies as careers progress (stability).

We study how these two affect the estimation of Q.

Accuracy. The Q-parameter is estimated based on sequences of data points with limited length. For this reason, variations of its estimation are inevitable. Indeed changes in the estimation of Q are even present in synthetic data produced by the model with fixed Q and finite N. This variance associated with the estimation of Q can be analytically determined. Indeed the Fisher information says that, for a scientist i, the uncertainty in the estimation of his/her Q_i parameter, $\sigma^2_{Q_i}$, is:

$$\Delta^2_{Q_i} = \left(-\frac{\mathrm{d}^2 \left(\log \mathcal{L}_i \right)}{\mathrm{d}Q_i^2} \right)^{-1/2} \tag{S31}$$

where $\mathcal{L}_i$ is the likelihood function defined in Eq. (S22). By plugging the expression of $\mathcal{L}_i$, we obtain:

$$\Delta^2_{Q_i} = \frac{\sigma_p}{\sqrt{N_i}}. \tag{S32}$$

For sufficiently large N_i (infinite sequences), the uncertainly converges to zero, as expected. We report in Fig. S31B the relative uncertainty on Q, $\Delta Q/Q$ as a function of the number published papers N predicted by the model for the APS dataset. This is always smaller than 20%, it drops below 10% after a scientist has published 50 papers and is less than 1% after 300 papers.

When we measure the same quantity for the data, reported in Fig. S31D, we see the same average trend as predicted by the model, but with a larger envelope (light blue area). These changes are due to the stability of the Q parameter as explained in the paragraph below.

Stability. Changes of the value of Q that are not explained by the number of data points used occur if Q is not a perfect parametrization of the impact of a career. That is, the changes observed in the values of Q measured at different career stages cannot be explained by the stochastic uncertainty of the model.

For the APS data, in 75% of the cases the estimated Q-parameter lies within the uncertainty envelope provided by the model – in these cases a variation in Q can be fully explained by the

inherent stochastic variance. In 25% of the cases the variation in Q is higher than the variation predicted by the stochastic nature of the model. However, given that the magnitude of this surplus variation never exceeds 15% and that the average relative error is always below 10%, the Q-parameter is estimated with a reasonably high accuracy (Fig. S32.

To test the stability of the Q parameter throughout the overall career, and not as a function of the number of papers considered N, we consider careers with at least 50 papers and calculate their early and late Q-parameter, Q_{early} and Q_{late} respectively, using Eq. S28 on the first and second half of their papers, respectively. In this case the stochastic uncertainty explains the differences between Q_{early} and Q_{late} for the large majority of scientists (95.1%, Fig. S31E).

Overall the error in estimation of Q is higher at the beginning of a career, being as high as 35% for the first 10 papers, but decreases fast as a function of N, dropping to below 20% for $N > 50$. We are confident that further research on the model and the application of advanced machine learning techniques will improve the accuracy of Q, making the predictions based on Q even more accurate and actionable.

S4.10 Predictions of the Q-model: highest impact paper

The Q-model, after the estimation of its parameters, allows to make predictions about the scaling of $\langle \log c_{10}^* \rangle$ with productivity N and with the logarithm of the average impact $\langle \log \bar{c}_{10}^* \rangle$.

The impact of the largest impact publication of a scientist, c_{10}^*, as a function of $\hat{Q}$ and productivity $\hat{N}$ is given by the sum of the highest potential impact s/he extracts, $\hat{p}^*$, summed to his/her $\hat{Q}$:

$$\log c_{10}^* \left(\hat{N}, \hat{Q}\right) = \hat{p}^* \left(\hat{N}, \hat{Q}\right) + \hat{Q}. \tag{S33}$$

The calculation of $\hat{p}^* \left(N\right)$ is analogous to the R-model calculation of c_{10}^* (Eq.(S8)), where instead of the probability distribution $P\left(c_{10}\right)$ we need to use the conditional probability $P\left(\hat{p}|\hat{Q}, \hat{N}\right)$. However, as proved in S4.4, we have that $P\left(\hat{p}|\hat{Q}, \hat{N}\right) = P\left(\hat{p}\right)$, meaning that $\hat{p}^*$ has no dependence on $\hat{Q}$, but only on $\hat{N}$, the latter dependence due purely to the extreme statistics prediction. Hence, when integrating over all possible value of $\hat{Q}$ and taking the expected value we obtain

$$\begin{aligned}
\langle \log c_{10}^* \rangle \left(\hat{N}\right) &= \int_0^\infty \left[\hat{Q} + \langle \hat{p}^* \rangle (\hat{N})\right] P(\hat{Q}|\hat{N}) \mathrm{d}\hat{Q} = \\
&= \int_0^\infty \hat{Q}\ P(\hat{Q}|\hat{N}) \mathrm{d}\hat{Q} + \langle \hat{p}^* \rangle (\hat{N}) = \\
&= \int_0^\infty \hat{Q}\ P(\hat{Q}|\hat{N}) \mathrm{d}\hat{Q} + e^{\hat{N}} \int_0^\infty \hat{p} P(\hat{p}) C(\hat{p})^{e^{\hat{N}}-1} \mathrm{d}\hat{p}
\end{aligned} \tag{S34}$$

where $C(\hat{p}) = \int_0^{\hat{p}} P(\hat{p}') \mathrm{d}\hat{p}'$. The first integral corresponds the average $\hat{Q}$ as a function of $\hat{N}$, which can be easily calculated if we know the form of $P(\hat{Q}|\hat{N})$. This distribution, being the conditional derived by a bivariate normal distribution, is a normal distribution with mean

$\bar{\mu} = \mu_Q + \frac{\sigma_{Q,N}}{\sigma_N^2}\left(\hat{N} - \mu_N\right)$ and variance $\bar{\sigma}^2 = \sigma_Q^2 - \frac{\sigma_{Q,N}^2}{\sigma_N^2}$ (*85*). The second integral in Eq. (S34), can be only calculated numerically. By substituting $\hat{N} = \log N$ we obtain

$$\langle \log c_{10}^* \rangle \left(\hat{N}\right) = N \int_0^\infty \hat{p} P\left(\hat{p}\right) C\left(\hat{p}\right)^{N-1} d\hat{p} + \frac{\sigma_{Q,N}}{\sigma_N^2} \log N + \mu_Q - \mu_N \frac{\sigma_{Q,N}}{\sigma_N^2}. \tag{S35}$$

To determine the relation between $\langle \log c_{10}^* \rangle$ and $\langle \log c_{10}^{-*} \rangle$, we make use of the observation that $\hat{Q} = \langle c_{10}^{-*} \rangle - \langle \hat{p} \rangle = \langle c_{10}^{-*} \rangle - \mu_p$, and again that $\log c_{10}^* = \hat{p}^* + \hat{Q}$. We then compute the expected value of the largest impact publication, $\langle \log c_{10}^* \rangle$ as a function of the Q-parameter

$$\begin{aligned}
\langle \log c_{10}^* \rangle (\hat{Q}) &= \hat{Q} + \int_0^\infty \hat{p}^*(\hat{N}) P(\hat{p}, \hat{N}|\hat{Q}) \, \mathrm{d}\hat{p} \, \mathrm{d}\hat{N} = \\
&= \langle c_{10}^{-*} \rangle - \mu_p + \int_0^\infty \hat{p}^*(\hat{N}) P(\hat{p}) P(\hat{Q}|\hat{N}) \, \mathrm{d}\hat{p} \, \mathrm{d}\hat{N} = \\
&= \langle c_{10}^{-*} \rangle - \mu_p + \int_0^\infty \hat{p}^*\left(N\right) P\left(\hat{N}|\langle \log c_{10}^{-*} \rangle - \mu_p\right) d\hat{N}
\end{aligned} \tag{S36}$$

where $\hat{p}^*\left(N\right) = \int_0^\infty \hat{p}' P\left(\hat{p}'\right) C\left(\hat{p}'\right)^{N-1} d\hat{p}'$, as calculated for Eqs. (S8) and (S34), while $P\left(\hat{N}|\langle \log c_{10,i}^{-*} \rangle - \mu_p\right)$ is the conditional probability obtained from the bivariate normal distribution $P(\hat{Q}, \hat{N})$, and has mean $\bar{\bar{\mu}} = \mu_N + \frac{\sigma_{Q,N}}{\sigma_Q^2}\left(\langle \log c_{10}^{-*} \rangle - \mu_p - \mu_Q\right)$ and variance $\bar{\bar{\sigma}}^2 = \sigma_N^2 - \frac{\sigma_{Q,N}^2}{\sigma_Q^2}$. Eq. (S36), same as Eq. (6) of the main text, can be evaluated numerically.

S4.11 Predictions of the Q model: impact indicators

The Q parameter comprises all the information about a scientist's individual impact distribution, $P\left(c_{10}\right)$. For this reason, it is possible to estimate the value of individual impact indicators by means of Q and productivity N.

h-index. The original definition of the h-index states that "a scientist has index h if h of his/her N papers have at least h citations each, and the other $(N - h)$ papers have no more than h citations each" *(12)*. We denote this index as h_{c_∞}, since it is based on all the citations that a paper will ever gather. The Q-model does not provide a direct prediction for h_{c_∞}, since this would require to incorporate information about the long term impact of papers *(6)*. However, the Q-model allows the prediction of $h_{c_{10}}$, the h-index based on the citations a paper has gathered after 10 years. Because the large majority of papers gathers most citations within two or three years after publication *(6, 45)*, for many papers $c_{10} \simeq c_\infty$. For this reason the measurements of $h_{c_{10}}$ and h_{c_∞} are highly correlated (Fig. S33A), indicating that $h_{c_{10}}$ can be used as proxy for h_{c_∞}.

For an author with parameter Q and N publications, we have

$$h_{c_{10}} = N C_> \left(h_{c_{10}}|Q, N\right), \tag{S37}$$

where $C_{>}\left(h_{c_{10}}|Q,N\right)=\int_{h_{c_{10}}}^{\infty}P(c_{10}|Q,N)\mathrm{d}c_{10}$ corresponds to the probability that one of the scientist's paper has more than $h_{c_{10}}$ citations. We take the logarithm of both members of Eq. (S37) and remember that $\log c_{10}=\log p+\log Q$, with $\log p$ uncorrelated from $\log Q$ and $\log N$ (S4.4) and following a normal distribution with parameters (μ_p,σ_p). We obtain

$$\log h_{c_{10}}-\log\Phi\left(\frac{\log h_{c_{10}}-\log Q-\mu_p}{\sigma_p}\right)=\log N \tag{S38}$$

where $\Phi(x)\equiv\int_x^{\infty}\phi(x)dx=\frac{1-\mathrm{erf}(x/\sqrt{2})}{2}$ is the complementary cumulative distribution of the standard normal distribution $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Eq. (S38) is an implicit equation of $h_{c_{10}}$ and has no explicit solution for finite N. However it can be solved numerically, providing a prediction that is in excellent agreement with the data for the h-index at the end scientific careers (Fig. S33B), for the expected value $\langle h_{c_{10}}(N)\rangle$ as a function of N for scientists with similar Q (Fig. S33C) or even when we analyze the $h_{c_{10}}$ dynamic in individual careers (Fig. S34A). The numerical solution in Fig. S33C also shows that for large N, $h_{c_{10}}\sim N^{\gamma}$, with γ depending on the Q parameter. Hence, for fixed Q, asymptotically $h_{c_{10}}$ is a proxy of productivity N.

Total number of citations. The Q-model provides an exact prediction for a scientist's $C_{tot,c_{10}}$, the sum of his/her papers' citations 10 years after publication ($C_{tot,c_{10}}=\sum_{\alpha=1}^{N}c_{10,\alpha}$). $C_{tot,c_{10}}$ correlates highly with $C_{tot,c_{\infty}}$, which is the total number of citations a scientist will ever gather (Fig. S33D). We have that

$$C_{tot,c_{10}}\simeq N\int_0^{\infty}c_{10}P\left(c_{10}|Q,N\right)\mathrm{d}c_{10}=NQ\int_0^{\infty}pP\left(p\right)\mathrm{d}p \tag{S39}$$

where we have used the fact that p and (Q,N) are uncorrelated to write the second equality. The prediction of Eq. (S39) is in excellent agreement with the data when we consider $C_{tot,c_{10}}$ at the end scientific careers (Fig. S33E), the expected value of the $\langle C_{tot,c_{10}}(N)\rangle$ as a function of N for scientists with similar Q (Fig. S33F) or even when we analyze the dynamic of $C_{tot,c_{10}}$ in individual careers (Fig. S34B). Eq. S39 also indicates that $C_{tot,c_{10}}$ grows linearly with N, as observed in Fig. S33F.

g-index. The Q-model provides also a prediction for the g-index, defined as "the (unique) largest number such that the top most cited g articles of a scientist received at least g^2 citations" (*86*). Similarly to the h-index, the Q-model can predict exactly $g_{c_{10}}$, the g-index calculated using the paper citations after 10 years, c_{10}. To derive $g_{c_{10}}$, we start from an equation similar to (S37), where we substitute $g_{c_{10}}$ on the left term and $g_{c_{10}}^2$ in the lower extreme of the integral. We obtain

$$\log g_{c_{10}}-\log\Phi\left(\frac{2\log g_{c_{10}}-\log Q-\mu_p}{\sigma_p}\right)=\log N \tag{S40}$$

where $\Phi(x)\equiv\int_x^{\infty}\phi(x)dx=\frac{1-\mathrm{erf}(x/\sqrt{2})}{2}$ is the complementary cumulative distribution of the standard normal distribution $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

S4.12 Predictive Power of the Q-parameter

Given that the Q-parameter can be estimated early in a career with reasonable accuracy (see S4.9), it can be used to make future predictions of a scientist's impact, with important implications for science and science policy. We scrutinize here the predictive power of the Q-parameter in determining dynamical impact indicators like the h-index.

Given an author with $\hat{Q}$ and N publications. Defining π_i is the probability for a paper written by this author to have i number citations for $i = 0, 1, 2 \ldots$, we have

$$\sum_{i=0}^{\infty} \pi_i = 1. \tag{S41}$$

We define

$$C_<(i) \equiv \Phi\left(\frac{\ln i - \hat{Q} - \mu_p}{\sigma_p}\right), \tag{S42}$$

where, as in S37 $\Phi(x) \equiv \int_{-\infty}^{x} \phi(x)\mathrm{d}x = \frac{1+\mathrm{erf}(x/\sqrt{2})}{2}$, with $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ being the normal distribution. Then we have $\pi_i = C_<(i+1) - C_<(i)$.

Given N papers, assuming n_i number of papers being cited at least i times. The probability of find $\{n_0, n_1, \ldots\}$ simply follows a multinomial distribution

$$P(n_0, n_1, \ldots) = N! \prod_{i=0}^{\infty} \frac{1}{n_i!} \pi_i^{n_i},$$

with $\sum_{i=0}^{\infty} n_i = N$,

Given an non-negative integer number r, the number of papers s with citations equal to or larger than r is simply $s = \sum_{i=r}^{\infty} n_i$. The probability distribution $\phi(s|r, N)$ of finding s given r can be calculated as

$$\phi(s|r, N) = \sum_{n_0, n_1 \ldots} \delta\left(\sum_{i=r}^{\infty} n_i - s\right) P(n_0, n_1, \ldots).$$

It can be shown easily that $\phi(s|r, N)$ follows a binomial distribution

$$\phi(s|r, N) = \frac{N!}{s!(N-s)!} \left(\sum_{i=0}^{r-1} \pi_i\right)^{N-s} \left(\sum_{i=r}^{\infty} \pi_i\right)^{s} = \frac{N!}{s!(N-s)!} C_<(r)^{N-s} (1 - C_<(r))^s.$$

The complementary cumulative distribution instead follows

$$\begin{aligned} \phi_>(s|r, N) &= \sum_{n_0, n_1 \ldots} \Theta\left(\sum_{i=r}^{\infty} n_i - s\right) P(n_0, n_1, \ldots) = \\ &= \sum_{i=s}^{N} \frac{N!}{i!(N-i)!} C_<(r)^{N-i} (1 - C_<(r))^i \end{aligned} \tag{S43}$$

where $\Theta(x) = 1$ if $x \geq 0$ and 0 for $x < 0$.

Given historical record of N_0 papers with their citations $c_1, c_2, \ldots, c_{N_0}$, we want to estimate the future value of the h-index together with an uncertainty envelope. To this end, we need to calculate the probability that the scientist has h-index equal to h, $P(h)$, after publishing N_1 papers. We start by observing that there are $N' \equiv N_1 - N_0$ new papers with undetermined citations. Again we start with the complementary distribution

$$P_>(h) = \sum_{n_0,n_1\ldots} \Theta\left(\sum_{i=h}^{\infty}(n_i + m_i) - h\right) P(n_0, n_1, \ldots),$$

where $\{n_0, n_1, \ldots\}$ and $\{m_0, m_1, \ldots\}$ denote papers published after and before N_0 respectively. Define $f(x) = \sum_{i=x}^{\infty} m_i$ the number of papers published before N_0 has citation equal or greater than x, we have

$$f(x) = \sum_{i=1}^{N_0} \Theta(c_i - x).$$

Then

$$P_>(h) = \sum_{n_0,n_1\ldots} \Theta\left(\sum_{i=h}^{\infty} n_i - (h - f(h))\right) P(n_0, n_1, \ldots),$$

Comparing with Eq. S43, we find that

$$P_>(h) = \phi_>(h - f(h)|h, N') = \sum_{i=h-f(h)}^{N'} \frac{N'!}{i!(N'-i)!} C_<(h)^{N'-i}(1 - C_<(h))^i,$$

where $h \geq f(h)$, and the h-index distribution follows $P(h) = P_>(h) - P_>(h+1)$.

Finally we can use this expression of $P(h)$ to calculate the expected value of the h-index, $h_{c_{10}}(N|c_1 \cdots c_{N_0})$ and its uncertainty σ numerically:

$$h_{c_{10}}(N|c_1 \cdots c_{N_0}) = \int hP(h)\mathrm{d}h \tag{S44}$$

$$\sigma^2 = \int h^2 P(h)\mathrm{d}h - \left(\int hP(h)\mathrm{d}h\right)^2 \tag{S45}$$

To explore the Q-parameter predictive power in the data, we have measured the Q parameter using only data in the first part of a career, and used it in Eqs. (S44)-(S45) obtaining a prediction with an uncertainty envelope for the h-index as a function of the productivity N. In Fig. S35, we show the prediction of the h-index for 4 scientists by using information only about the first $N_0 = 20$ papers and, for a more accurate prediction, about the first $N_0 = 50$ papers. For $N_0 = 20$, the prediction of the h-index at $N = 40$ publication is accurate for 74% of the scientists, while for $N_0 = 50$ the accuracy of the prediction 20 papers later increases to 81%.

The early estimation of the Q-parameter is accurate by using only the first 20 papers of a scientist and provides a long-term prediction for impact metrics, like the h-index, which is in excellent agreement with the data even at $N = 160$, an 8-fold number of papers later in the career. Similar results can be obtained to predict the future total number of citations C_{tot}. Taken together, the Q-model provides long-term predictive power of career impact.

S5 Relation between the Q-parameter and productivity

The measured parameters of the Q-model on the APS dataset, reported in Eq. (2) of the main text, indicated that the Q-parameter and productivity N are correlated. The strength of the correlation is $\rho_{Q,N} = 0.34$, where we used the relation $\sigma_{Q,N} = \rho_{Q,N}\sigma_Q\sigma_N$. So, while it is true that very high impact work are the result of drawing a high p (luck) and high Q, this is more likely to happen to scientists with high productivity N, because of the non-zero $\rho_{Q,N} = 0.34$. However, being $\sigma_{p,Q} \simeq 0$ and $\sigma_{p,N} \simeq 0$, neither productivity nor Q have an effect on the luck component.

The relationship between Q and N is also present when we measure impact by rescaling according to the author list order (S6.2 and Fig. S40) and with the credit assignment algorithm (Figs. S43 and S42). Also, hints of this relationship are present in Fig. S24, where low max impact scientists do not produce enough papers to be included in the analysis for $80 \leq N \leq 100$.

In Fig. S49, we also points to a relation between early impact, quantified with $\langle c_{10} \rangle$ averaged over the first 10 papers in a scientist's career, and longevity, quantified with the overall productivity in a scientist's career. The effect is different from that found in randomized career, and is more pronounced for scientists that have a larger productivity N.

The random impact rule as well as the core results of the Q-model are however not affected by this coupling. Indeed, the random impact rule holds when we control from productivity (see Fig. S24), and all Q-model predictions are based on the uncoupling between p and (Q, N), but no assumption is made on the uncoupling between Q and N. Also, the fact that higher early impact is associated with higher productivity is consistent with the non-zero $\sigma_{Q,N}$. In this work we do not investigate where this association stems from. This calls for more future work on models of the co-evolution of *short-term* impact and productivity, which can explain this coupling.

S6 Controlling for coauthorship effects

Teams play an important role in the impact of coauthored publications *(32, 87–89)* affecting the career of an individual as well as his/her scientific impact. While the goal of this work is not to uncover the precise role of teams and collaboration on individual impact, here we test whether or not our main results are an artifact of team and collaborative effects, as discussed in the subsections below.

S6.1 Effect on the randomness of the highest impact paper

Two kinds of test are considered to check whether our findings are impacted by collaborative effects. In the first one, we select only scientists whose most cited paper is single-authored, where the collaborative effect does not play a role. In the second one, we distribute a paper impact among authors in the case of multi-authored papers, using three different credit allocation approaches: (i) each author gets the same credit, which is the approach we use in the manuscript; (ii) the credit share of an author is defined by her rank in the author list of the paper; (iii) credit is distributed among the authors based on the collective perception of the scientific community.

Solo authors. To test if impact could be a simple consequence of co-authorship with a more experienced or visible individual, or some other consequence of team formation, we tested our main findings for scientists whose highest impact paper is single-authored, amounting in total to 238 physicists in our dataset. We then repeated our of Fig. 1E-F of the main paper (Fig. S17). Although all these measurements are noisier due to smaller number of careers with single authored highest impact work, we find no qualitative difference with our previous results, suggesting that our findings are robust to co-authorship effects.

Credit share: author order allocation. Most current approaches to assign credit to the authors of a paper are based on the their order in the paper author list. We implemented the method proposed by *(36)* that associates the highest weight to the first and last author, while the second author has half their weight, the third author one third of their weight, and so on. Based on this method, we assign to author i of paper α a credit share $\tilde{q}_i^\alpha$. We then convert the originally paper impact $c_{10,i\alpha}$ by using the following rule: each paper of author i is associated with a fraction $c_{10,i\alpha}^{ord}$ of the original impact, that is $c_{10,i\alpha}^{ord} = c_{10,i\alpha} q_i^\alpha$. We repeated our measurements of Fig. 1E-F and Fig. 2C-D of the main paper and find that all our empirical results hold even if impact is rescaled according to the credit share of each author (Figs. S40).

Credit share: collective allocation. Recently, a framework has been proposed to automatically allocate credits based on the community perception *(22)*. The leading hypothesis of this methodology is that the information about the informal credit allocation within science is encoded in the detailed citation pattern of the respective paper and other papers published by the same authors on the same subject. Indeed, each citing paper expresses its perception of the scientific impact of a paper's coauthors by citing other contributions by them, conveying implicit information about the perceived contribution of each author.

Consider a paper p_0 with m coauthors $\{a_i\}$ $(1 \leq i \leq m)$. To determine the credit share of each author, they first identify all papers that cite p_0, forming a set $\mathcal{D} \equiv \{d_1, d_2, \cdots, d_l\}$. Next they identify all co-cited papers $\mathcal{P} \equiv \{p_0, p_1, \cdots, p_n\}$, representing the complete set of papers cited by papers in the set $\mathcal{D}$. The relevance of each co-cited paper p_j $(0 \leq j \leq n)$ to the target paper p_0 is characterized by its co-citation strength s_j between p_0 and p_j, defined as the number of times p_0 and p_j are cited together by the papers in $\mathcal{D}$. For example, for p_1 in Fig. S41A we

have $s_1 = 1$ since only one paper (d_1) cites p_0 and p_1 together, while $s_2 = 4$ as four papers (d_1, d_2, d_3, d_5) cite p_0 and p_2 together. Co-citation strength captures the intuition that papers by an author that are perceived to be very relevant to paper p_0 should increase the author's perceived contribution to p_0. Note that the target paper p_0 is also viewed as a co-cited paper of itself with co-citation strength equal to the citation count of p_0. Consequently for papers with high citation count the credit share of coauthors is less likely to be affected by other co-cited papers.

Using the author list of the co-cited papers, a credit allocation matrix $\mathbf{A}$ is calculated, whose element A_{ij} denotes the amount of credit that author a_i gets from co-cited paper p_j. The fractional credit allocation matrix does not depend on the order of authors in the author list. The total credit k_i of author a_i is the weighted sum of its local credit obtained from all co-cited papers

$$k_i = \sum_j A_{ij} s_j$$

or in the matrix form

$$\mathbf{k} = \mathbf{A}\mathbf{s} \tag{S46}$$

The vector $\mathbf{k}$ provides the credit of all authors of target paper p_0. By normalizing $\mathbf{k}$ the fractional credit share among coauthors is obtained (Fig. S41e).

Based on this methodology, we assign to author i of paper α a credit share $\tilde{k}_i^\alpha$. We then convert the originally paper impact $c_{10,i\alpha}$ into two different ways: (i) each paper of author i is associated with a fraction $c_{10,i\alpha}^{share}$ of the original impact, that is $c_{10,i\alpha}^{share} = c_{10,i\alpha} k_i^\alpha$, (ii) each paper of author i is associated an impact $c_{10,i\alpha}^{max}$ equal to its original impact $c_{10,i\alpha}$ only if author i has the maximum credit share k_i^α among all the authors of α, otherwise $c_{10,i\alpha}^{max} = 0$. We repeated our measurements of Fig. 1E-F and Fig. 2C-D of the main paper and find that all our empirical results hold even if impact is rescaled according to the credit share of its author (Figs. S42-S43). This indicates that the presence of different teams in an individual's career and the resulting different paper impact share does not affect the random impact rule, the necessary assumption for our modelling framework.

S6.2 Effect on the Q parameter

We studied the robustness of the Q parameter against the removal of coauthored papers. More specifically, to test if one particular coauthor could skew the Q parameter, we computed the variation of the Q-parameter for a scientist i, Q_i^{-j}/Q_i by using only the subset of publications of scientist i that lack all papers written with co-author j. We find that the $P\left(\log Q_i^{-j}/Q_i\right)$ distribution is peaked is at zero (Fig. S44) and decays approximately exponentially for high Q_i^{-j}/Q_i. Therefore in most cases the removal of specific coauthors (and the papers written with them) does not alter Q and when it does, the effect remains small, as indicated by the exponential tail. Furthermore, the observed changes do not exhibit a bias towards higher or lower Q (symmetric distribution in respect to zero). These properties hold when considering different values of the individual Q parameter (different colors in the plot below).

S7 Receiving Operating Characteristic (ROC) curve

A ROC curve measures the performance of a binary classifier system when its discrimination threshold T is varied (*89*). It is created by plotting for each T the true positive rate (TPR, true positive over all the positive, *e.g.* number of Nobel Laureates detected within the rank threshold T over all Nobel Laureates present in the dataset) vs the false positive rate (FPR, false positive over all the negative, *e.g.* number of non-Laureates detected within the threshold T over all non-Laureates present in the dataset). For example, let us assume we have a pool of 200 scientists, composed of 25 Nobel Laureates and 175 non-Laureates. If we consider the first 100 rank positions of a ranking and find 5 Nobel Laureates out of 25 present in the dataset, and 95 non-Laureates out of 175 non-Laureates scientists, we have T=100, $TPR = 5/25 = 20\%$ and $FPR = 95/190 \simeq 54.3\%$. When $T = 200$, we consider the entire pool of scientists in the dataset and by definition we have $TPR = 100\%$ and $FPR = 100\%$. Therefore, a ROC-plot is a curve connecting the points $(0, 0)$ and $(1, 1)$ in the so-called ROC space, where FPR is reported in the x-axis and TPR in the y-axis and both quantities by definition vary between 0 and 1. The best possible ranking yields a ROC-curve that from $(0, 0)$ jumps immediately to the point of coordinates $(0, 1)$ (upper left corner) and is constant afterwards up to the point $(1, 1)$. A completely random guess instead yields a diagonal line, called also line of no-discrimination. A ranking yielding a ROC-curve placed in the upper diagonal part of the ROC-space performs better than a random ranking, while a ROC-curve in the bottom diagonal part performs worse than a random guess. A more precise way to quantity the overall performance of a ranking is based on measuring its *accuracy*, that is the area contained under the ROC curve. The maximum of accuracy, corresponding to the best ranking, is 1, while a random ranking has accuracy 0.5. For each rank threshold we also measure the *recall*, which simply corresponds to the TPR (in our case, the fraction of the total Nobel laureates detected within the threshold), and the *precision*, that is the number of true positive divided by the considered rank threshold (in our case, the fraction of scientists that are Nobel laureates within the rank threshold considered). A perfect ranking has a recall of 100%, corresponding to no false negatives, for all rank thresholds. It has also the best precision, that is no false positives, for all rank thresholds (*89*). In Fig. S45 we report recall and precision as a function of different rank thresholds for all rankings of Nobel laureates shown in Fig. 3D of the main text.

ROC-plots of Boltzmann and Dirac medalists are instead reported in Fig. S46, while their recall and precision is shown in Fig. S47.

S7.1 Early prediction of Nobel Laureates

We also test the ability of the early estimation of Q to predict future Nobel Laureates, in respect to their early productivity, citations, h-index and highest cited paper. In order to do this, we have first identified all Nobel Laureates that won the Nobel prize late in their career ($\sim$80% of them) and measured Q in the first, 6, 10 and 15 years of activity, as well as their h-index, highest impact paper c^*_{10}, total number of citations C_{tot} and of papers N. We then used these

measurements to construct ROC-plots (Fig. S48). The Nobel Prize winners are still placed at the top of the ranking, and the also the early Q-parameter performs better than all other indicators.

S8 Supplementary Figures

All results in the following refer to the APS dataset, unless otherwise noted.

S8.1 Characterization of the datasets

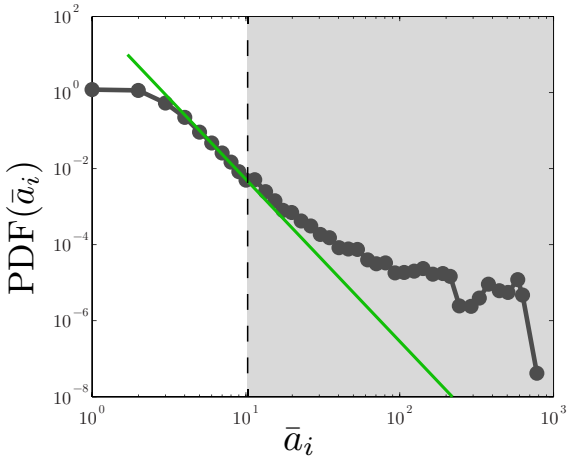

Figure S1: **Distribution of number of authors per paper $\bar{a}_i$.** For each paper i in the dataset, we denote with $\bar{a}_i$ the number of its authors and report the distribution $P(\bar{a}_i)$. The vertical line falls at ten authors, corresponding roughly to the point were the distribution deviates from the power law fitting line. We thus do not take into account in the disambiguation process all the papers that have more than 10 authors, corresponding to only 3% of all the papers.

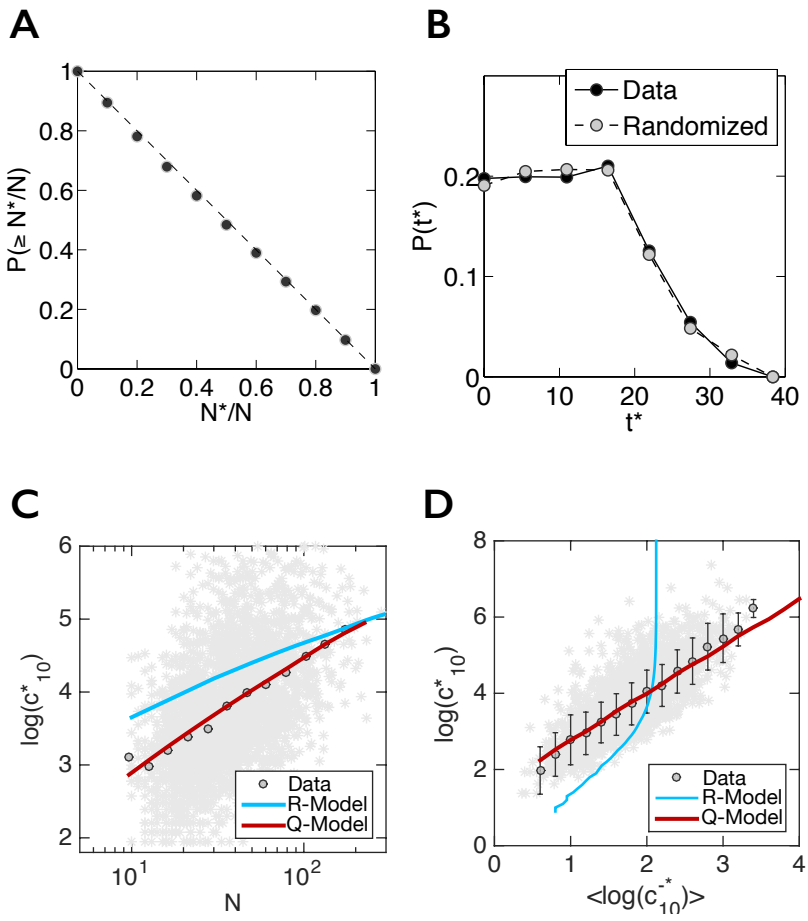

Figure S2: **Randomness of the highest impact paper and Q-model predictions for the subset of scientists with uncommon family names.** For the subset of scientists that do not have a common last name (see S1.1) we repeat the following measurements: **(A)** Cumulative distribution $P\left(\geq N^*/N\right)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of papers published by a scientist; **(B)** Distribution of the publication time t^* of the highest impact paper c_{10}^* for scientists' careers (black circles) and for randomized impact careers (grey circles). **(C)** Citations of the highest impact paper, c_{10}^*, vs the number of publications N during a scientist's career. Circles correspond to the logarithmic binning of scattered data, cyan curve corresponds to the prediction of the R-Model model and the red curve corresponds to the analytical prediction Eq. (S35) of the Q-Model model; **(D)** $\log c_{10}^*$ vs $\langle \log c_{10}^{-*} \rangle$, where $\langle \log c_{10}^{-*} \rangle$ is the average logarithm of the impact of a scientist's papers excluding the highest impact paper c_{10}^*. We report in cyan the R-Model prediction and in red the analytical prediction Eq. (S36) of the Q-model. All the measurements show that our findings remain unchanged when removing the careers of scientists with common asian names, indicating that our results are robust to the unavoidable small errors induced by the disambiguation procedure.

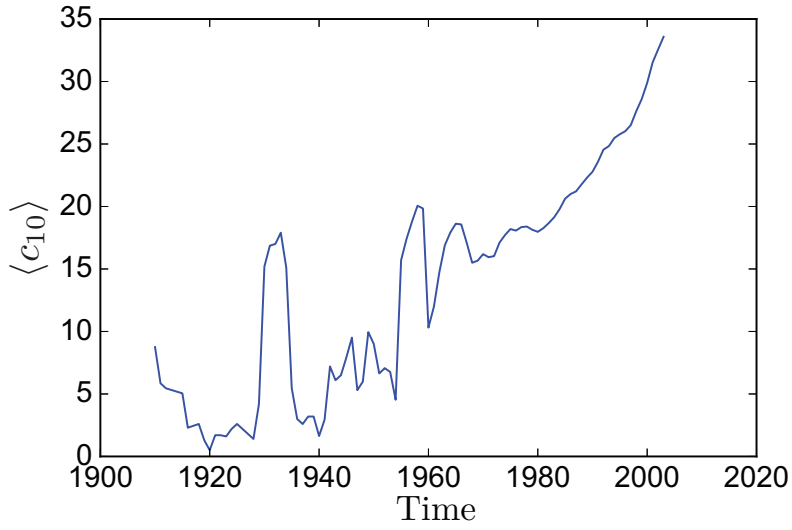

Figure S3: **Average number of citations $\langle c_{10} \rangle$ over time.** For the WoS dataset, the average impact of papers, quantified with number of citations 10 years after publication, steadily grows as a function of the publication year. For this reason, in the WoS data we rescale the raw number of citations c_{10} to correct for this citation inflation.

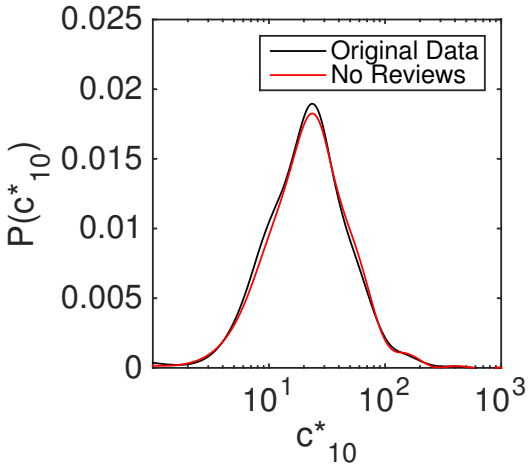

Figure S4: **Distribution of most cited paper.** We report the distribution of most cited paper $P\left(c_{10}^{*}\right)$ in the original dataset (black line) and in the dataset where we remove the papers published in Review of Modern Physics (red line). The two distributions are almost perfectly overlapping, showing that only in rare cases review papers are the most cited paper.

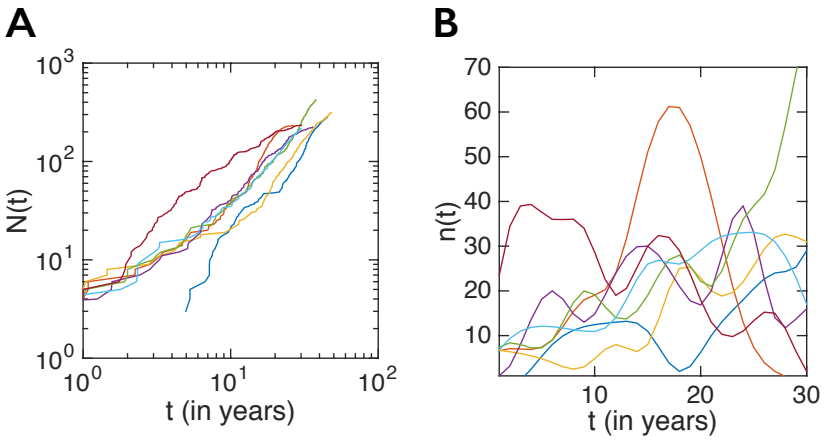

Figure S5: **Individual productivity over time. (A)** The number of papers $N\,(t)$ up to time t for scientists with at least 200 publication. The paper publication date is known with a time resolution of days; $t\,=\,0$ coincides with the day of the scientist's first published paper. Each curve can be asymptotically fitted with $N\,(t)\,\sim\,t^{\gamma}$, as indicated in Eq. (1) *(18)*.For each scientist, we extract the exponent γ based on the cumulative productivity in the second half of his/her career paper sequence. (B) Number of papers $n\,(t)$ published at time t for the same scientists in (A). $n\,(t)$ is the derivative of $N(t)$, hence following the equation $n\,(t)\,\sim\,t^{\gamma-1}$. At odds with $N(t)$, the noisy nature of $n(t)$ does not offer the possibility to fit the exponent γ.

S8.2 Impact trends before and after the highest impact work

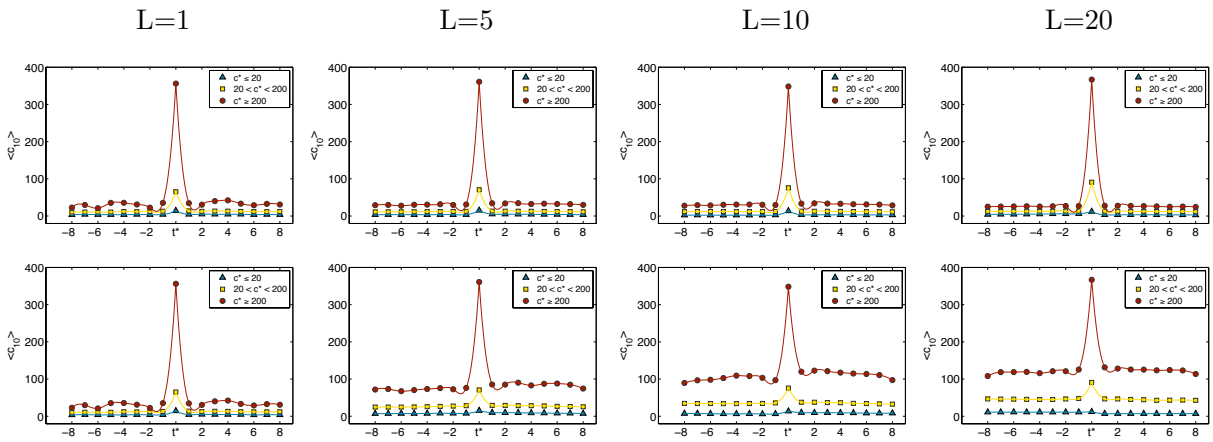

Figure S6: **Finger curves on smoothed careers.** We smooth the scientists' career by using a moving average (top panels) and a moving record (bottom panels) as described in S2.1, for different length of windows L. When $L = 1$ the careers are not smoothed and the finger curve is computed directly on the original data. For various values of L in both techniques no qualitative difference is observed, as no patterns preceding and following the highest impact paper appear.

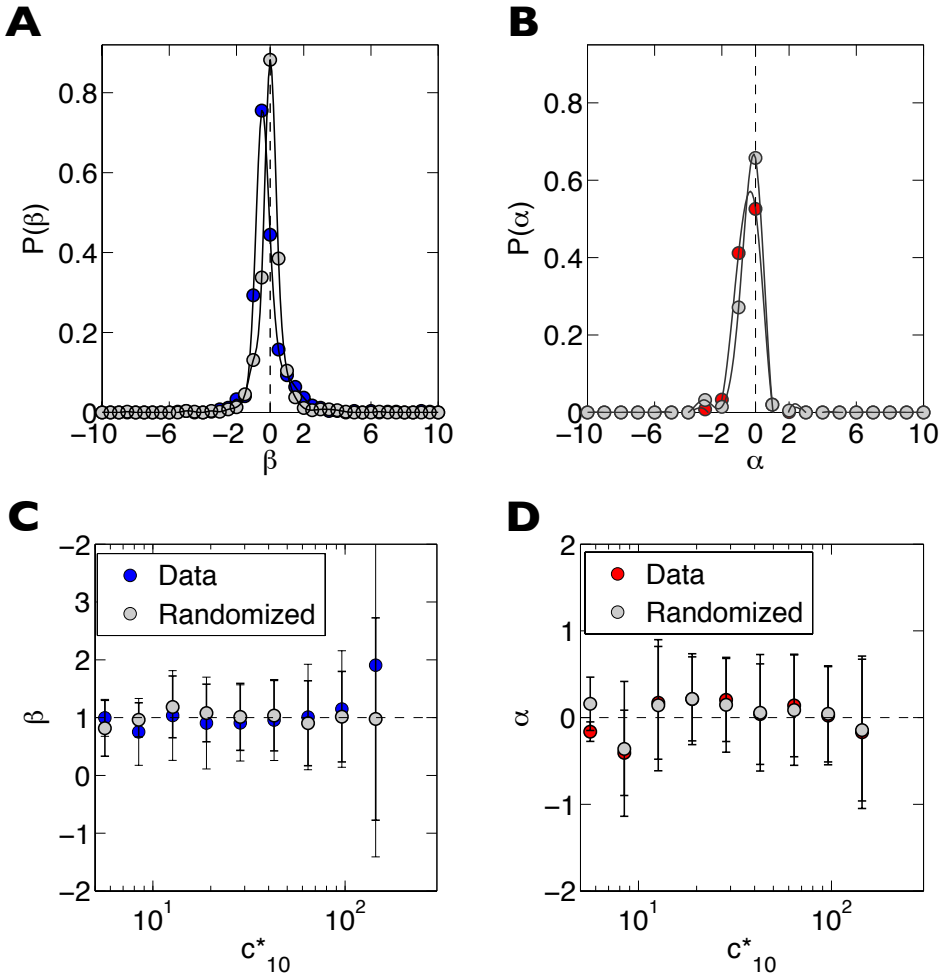

Figure S7: **Trends before and after the highest impact paper.** In **(A)** and in **(B)** we show the distribution of β and α respectively, provided by fitting Eqs. (S2) and (S3) to individual careers before and after the timing of the highest impact paper t^*. In the bottom panels, we report the average value and standard deviation of **(C)** β and **(D)** α for groups of scientists with similar c^*_{10}. The blue and red circles correspond to plots based on the original data, while grey circles correspond to randomized careers. In the randomized data the time of each publication and c^*_{10}, the impact of the highest impact paper, is preserved while the impact of all other papers is shuffled within the career. Mann-Whitney U tests confirm that the distributions based on data and randomized careers are indistinguishable ($p_{value} = 0.15$ and $p_{value} = 0.34$ for $P(\beta)$ and $P(\alpha)$ respectively).

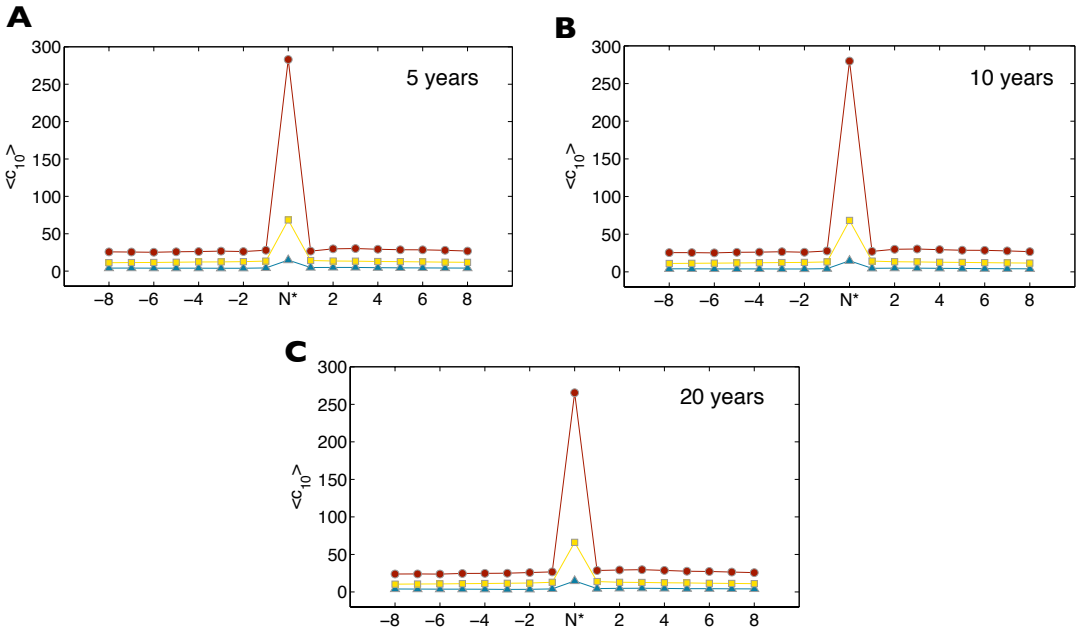

Figure S8: **Average impact before and after the highest impact paper for careers of different length.** The average impact $\langle c_{10} \rangle$ of papers published before and after the highest impact paper, c_{10}^* for high, middle and low impact scientists having at least **(A)** a 5-year long career, **(B)** a 10-year long career, **(C)** a 20-year long career (dataset analyzed in the manuscript). All plots document a lack of patterns before and after the highest impact paper.

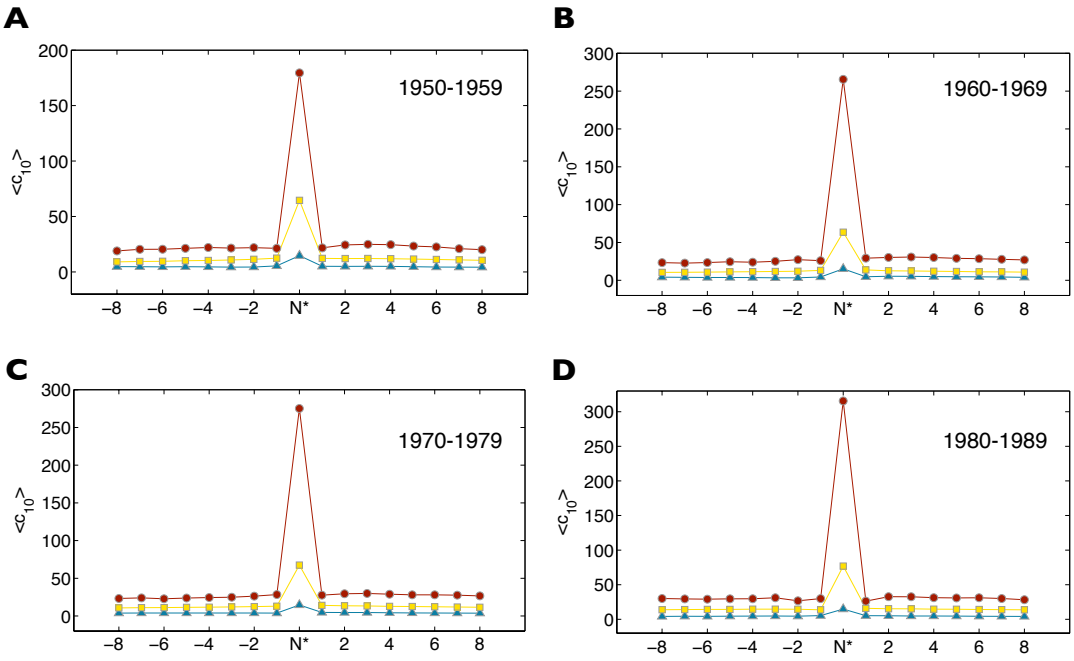

Figure S9: **Average impact before and after the highest impact paper for careers starting in different decades.** The average impact $\langle c_{10} \rangle$ of papers published before and after the highest impact paper, c_{10}^{*} for high, middle and low impact scientists starting their career, at least 10 years long, **(A)** between 1950 and 1959 (732 scientists), **(B)** between 1960 and 1969 (1,766 scientists), **(C)** between 1970 and 1979 (2,666 scientists), **(D)** between 1980 and 1989 (3,953 scientists). All plots document a lack of patterns before and after the highest impact paper.

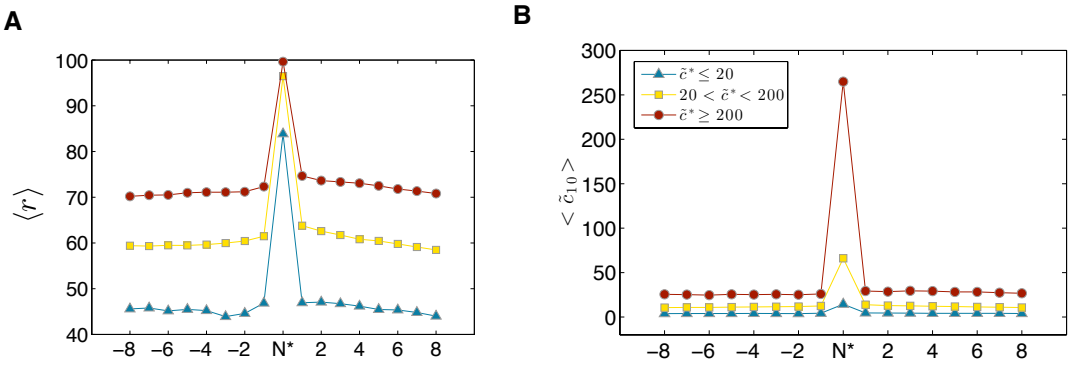

Figure S10: Average impact before and after the highest impact paper based on **(A)** rank and **(B)** time rescaled impact (see definitions in S1.6).

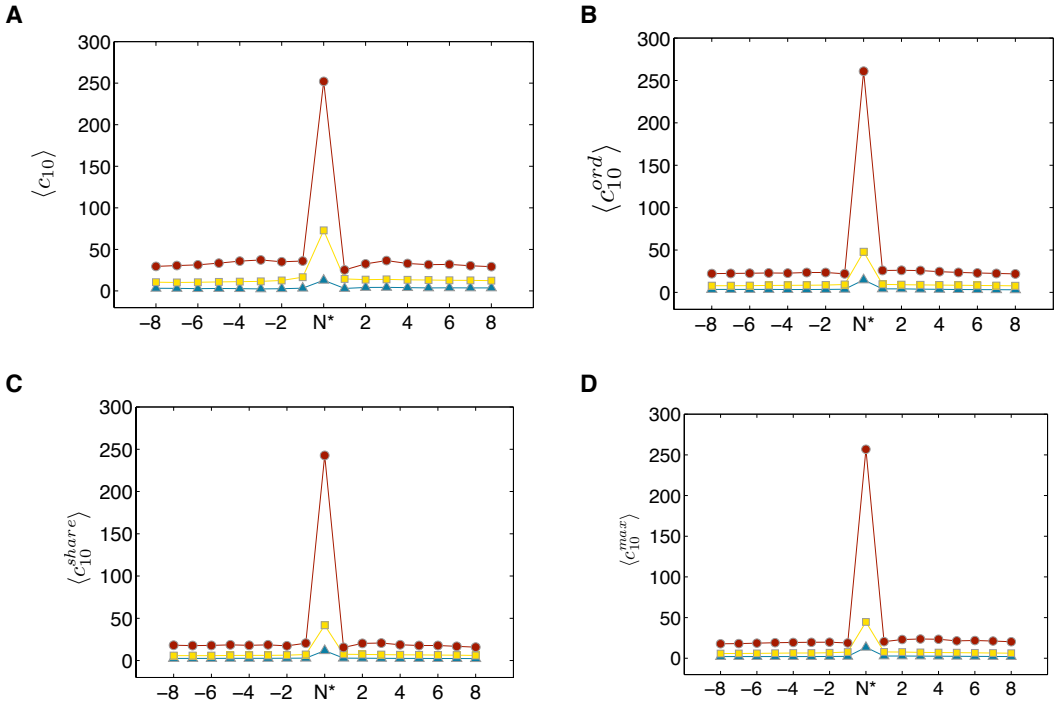

Figure S11: **Average impact before and after the highest impact paper controlling for co-authorship effects. (A)** Dataset with scientists that are the sole author of their highest impact paper, **(B)** dataset where impact is assigned based on the author list order, **(C)** dataset where impact is assigned proportionally to the credit share, **(D)** dataset where impact is assigned based on maximum credit share. See S6 for details on the datasets.

S8.3 Random Impact Rule

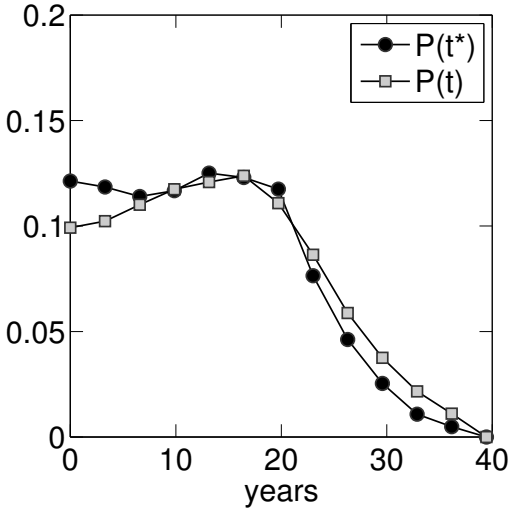

Figure S12: **Timing of the highest impact paper vs time of all papers.** Distribution of the publication time t^* of the highest impact paper during a scientist's career compared to the probability $P(t)$ of publishing a paper of any impact at time t. Both distributions are computed considering all careers of the subset with sustained productivity, analyzed in the main text. The similarity between the two curves further confirms the random impact rule, indicating that the highest impact paper can be any of a scientist's paper. The drop of $P(t^*)$ after 20 years is simply explained by the decreased probability of publishing, as a consequence of selecting scientists with a career of at least 20 years.

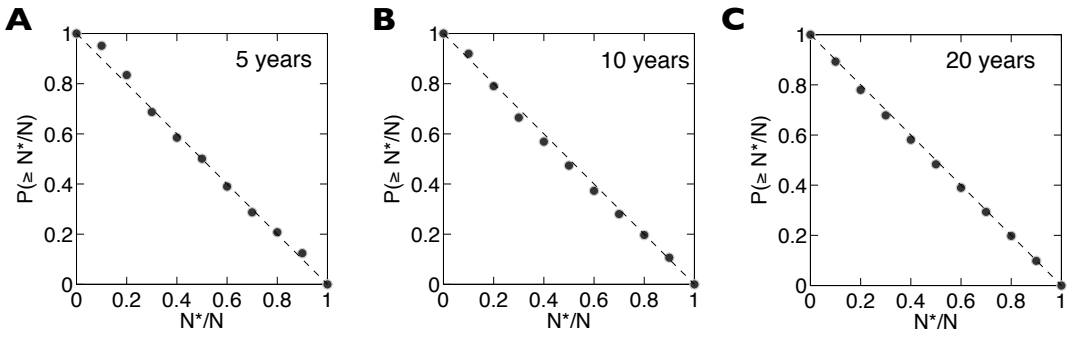

Figure S13: **Cumulative distribution $P(\geq N^*/N)$ for careers starting of different length.** Cumulative distribution $P(\geq N^*/N)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career, for scientists having at least **(A)** a 5-year long career, **(B)** a 10-year long career, **(C)** a 20-year long career (dataset analyzed in the manuscript). In all three cases the cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of papers published by a scientist.

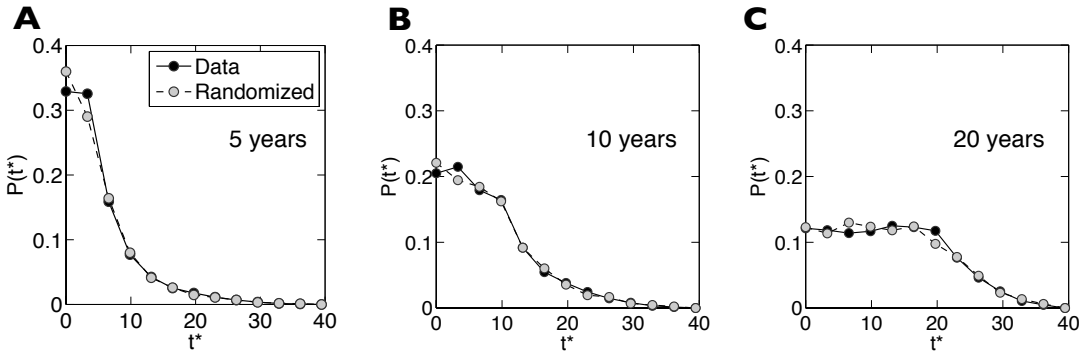

Figure S14: **Timing of the highest impact paper.** Distribution of the publication time t^* of the highest impact paper c_{10}^* for scientists' careers (black circles) and for the corresponding randomized impact careers (grey circles) for individuals having at least **(A)** a 5-year long career, **(B)** a 10-year long career, **(C)** a 20-year long career (dataset analyzed in the manuscript). While the specific shape of the distribution $P\left(t^*\right)$ is different for each subset, partially due to the constraint on the career length, the lack of differences between the curve corresponding to the data and the one corresponding to randomized careers confirms that the random impact rule holds regardless of the specific mechanisms shifting the peak of $P\left(t^*\right)$.

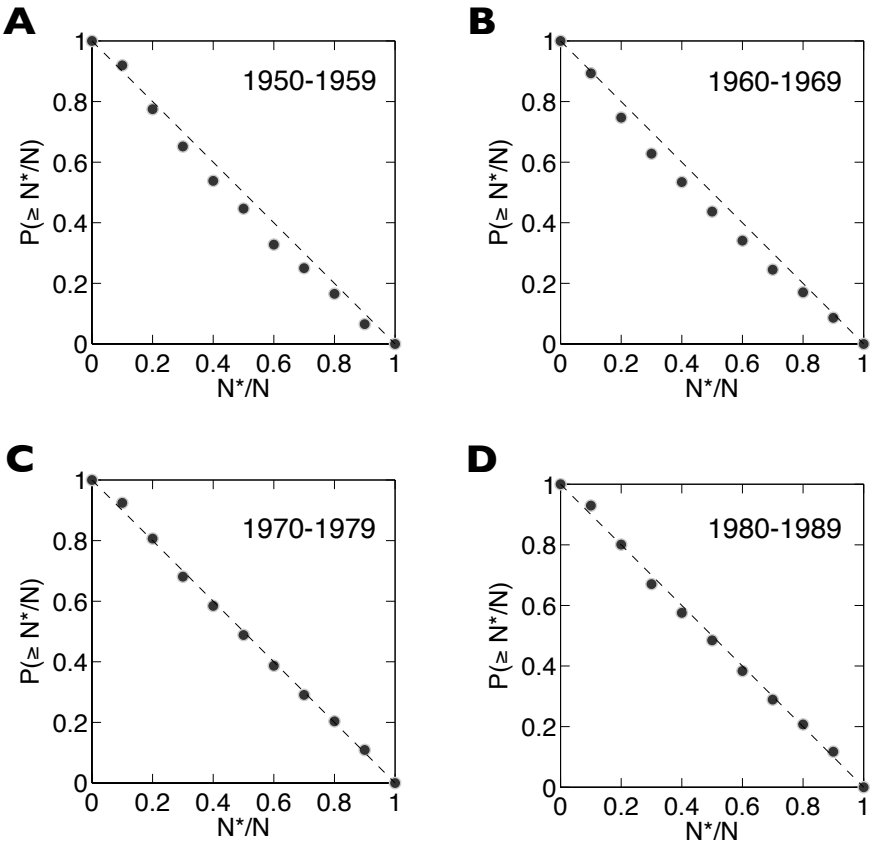

Figure S15: **Cumulative distribution $P(\geq N^*/N)$ for careers starting in different decades.** Cumulative distribution $P(\geq N^*/N)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career, for scientists starting **(A)** between 1950 and 1959, **(B)** between 1960 and 1969, **(C)** between 1970 and 1979, **(D)** between 1980 and 1989. In all four cases the cumulative distribution of N^*/N is a straight line with slope -1, indicating that $N*$ has the same probability to occur anywhere in the sequence of papers published by a scientist.

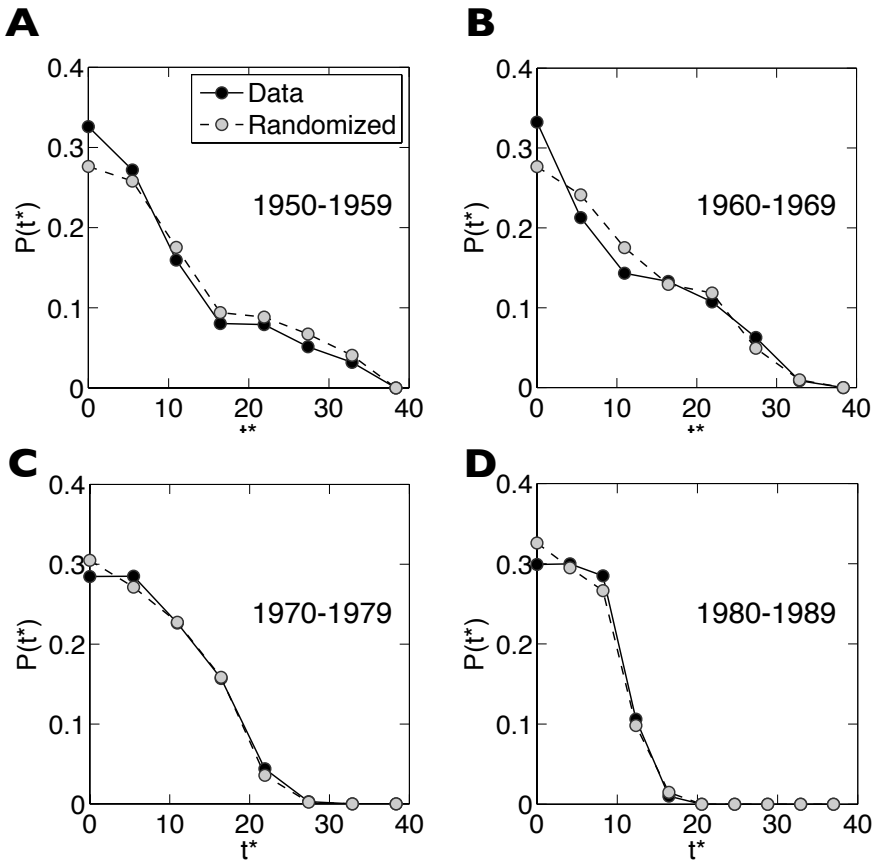

Figure S16: **Timing of the highest impact paper.** Distribution of the publication time t^* of the highest impact paper c_{10}^* for scientists' careers (black circles) and for the corresponding randomized impact careers (grey circles) for individuals starting at least 10 years long career, **(A)** between 1950 and 1959, **(B)** between 1960 and 1969, **(C)** between 1970 and 1979, **(D)** between 1980 and 1989. While the specific shape of the distribution $P(t^*)$ changes over decades, the lack of differences between the curve corresponding to the data and the one corresponding to randomized careers confirms that the random impact rule holds regardless of the specific mechanisms shifting the peak of $P(t^*)$.

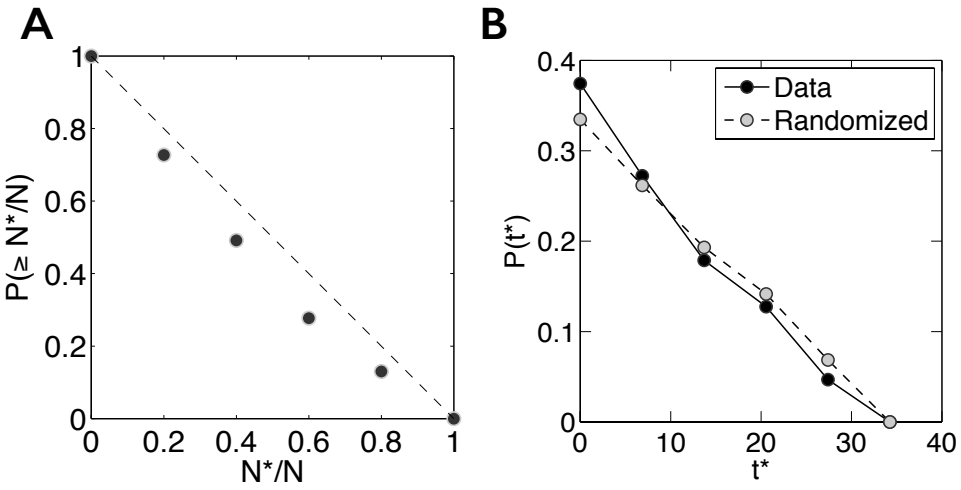

Figure S17: **Randomness of the c^*_{10} for scientists that are the sole author of their highest impact paper.** For the subset of scientists who are the single author of their highest impact paper we repeat the following measurements: **(A)** Cumulative distribution $P\left(\geq N^*/N\right)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career (Fig. 2E in the main text)). **(B)** Distribution of the publication time t^* of the highest impact paper c^*_{10} for scientists' careers (black circles) and for randomized impact careers (grey circles). The measurements show that the random impact rule is not an artifact induced by paper coauthorship.

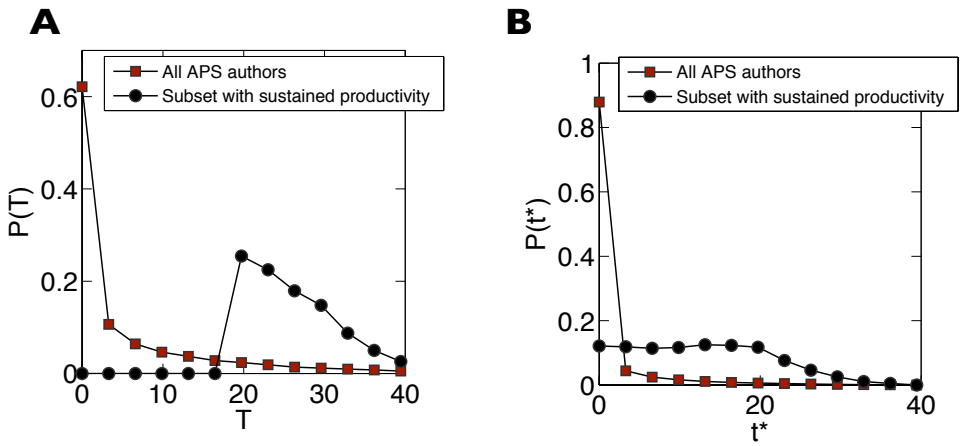

Figure S18: **Longevity and timing of the highest impact paper for all APS authors and authors with sustained productivity.** **(A)** Distribution of longevity T, i.e. time between a scientist's first and last publication, and **(B)** distribution of the publication time t^* of the highest impact paper in a scientist's career, for the full APS dataset, containing all individuals that have authored at least one paper in one of the APS journals (red squares) and for the authors with sustained productivity, studied in the main text (black circles). The authors with sustained productivity are those who (i) have authored at least one paper every 5 years, (ii) have published at least 10 papers, (iii) their publication career spans at least 20 years, as described in section S1.3. When considering all authors, the timing of the highest impact paper peaks very early in the career, a simple consequence of the short longevity of most authors, that publish only one or a few papers before leaving science.

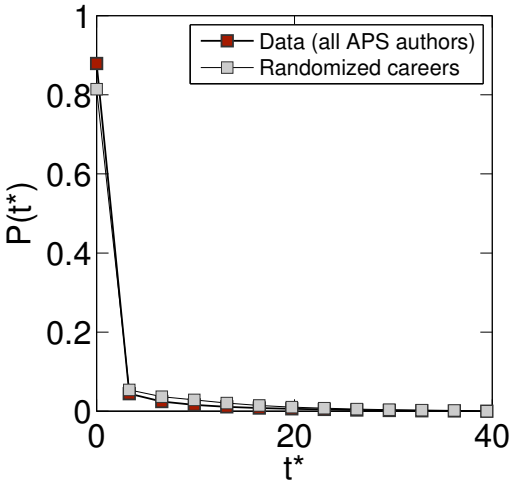

Figure S19: **Timing of the highest impact paper for all APS authors.** Distribution of the publication time t^* of the highest impact paper for scientists' careers (red squares) and for randomized impact careers (grey squares). The lack of differences between the two curves further confirms the random impact rule also for authors that do not have sustained productivity and/or high longevity.

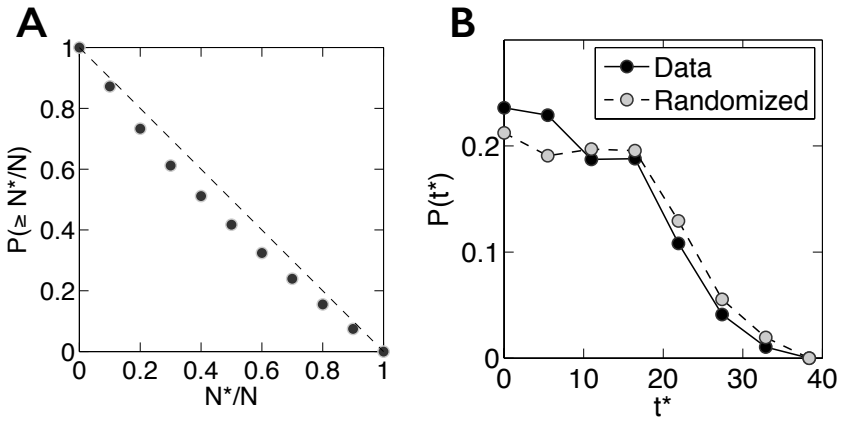

Figure S20: **Empirical observations for rank-based impact.** We associate each paper with an impact r based on its c_{10}-rank among all the papers published in the same 1-year window. A paper with $r = 100$ has the highest c_{10} in the 1-year window, while a paper with, say, $r = 60$ is ranked at the top 40% for c_{10} in the corresponding 1-year window. For our subset of scientists, we find that the top 27% scientist has the highest rank paper in the top 1% ($r^* \geq 99$, considered as high impact), the middle 58% (middle impact) has the highest rank paper between the top 1% and the top 10% ($90 < r^* < 99$, middle impact), while the remaining 15% of scientists has the highest rank paper in the bottom 90% ($r^* \leq 90$, low impact). Even the scientist at the bottom of the ranking for highest rank papers has $r^* = 49.7$, meaning that his/her highest achievement is median among all papers in a 1-year window. **(A)** Cumulative distribution $P(\geq N^*/N)$, where N^*/N denotes the order N^* of the highest rank paper in a scientist's career, varying between 1/N and 1. **(B)** Distribution of the publication time t^* of the highest impact paper r^* for scientists' careers (black circles) and for randomized impact careers (grey circles).

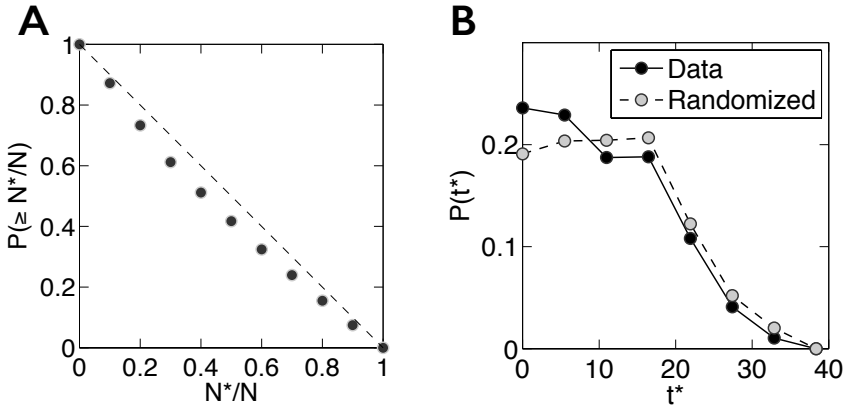

Figure S21: **Empirical observations for rank-based impact.** Same as in Fig. S20, where we consider a 5-year window instead. **(A)** Cumulative distribution $P(\geq N^*/N)$, where N^*/N denotes the order N^* of the highest rank paper in a scientist's career, varying between 1/N and 1. **(B)** Distribution of the publication time t^* of the highest impact paper r^* for scientists' careers (black circles) and for randomized impact careers (grey circles).

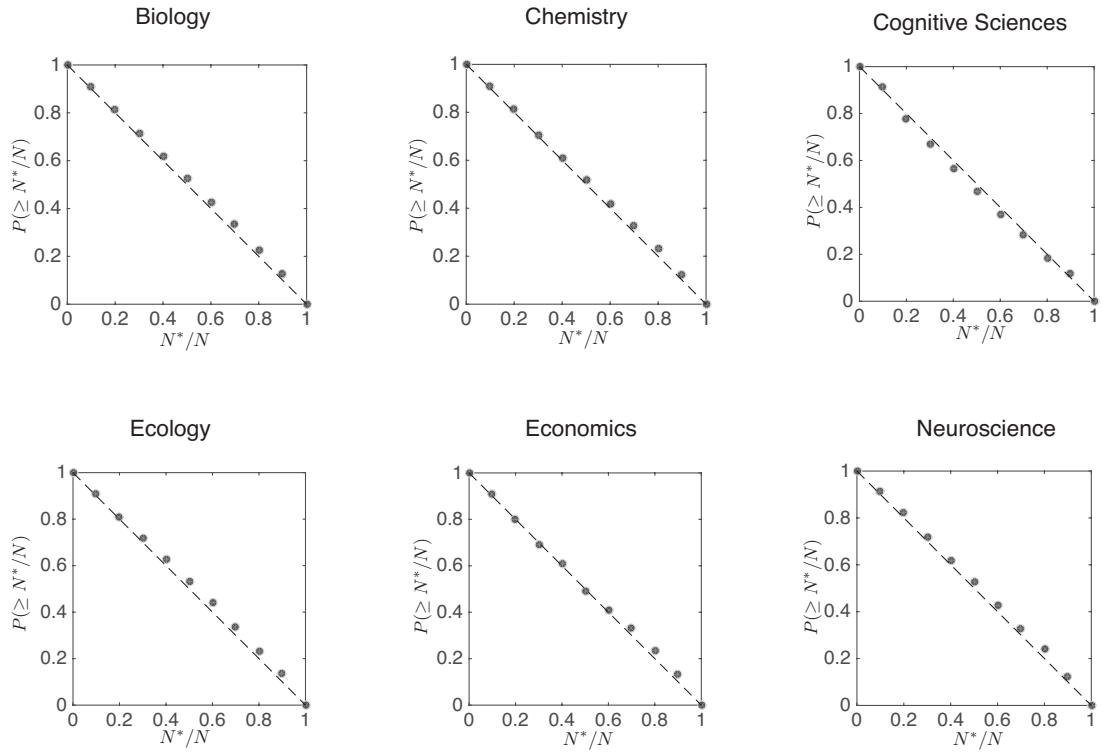

Figure S22: **Timing of the highest impact paper in the six different disciplines of the WoS dataset**. Distribution of the publication time t^* of the highest impact paper for scientists' careers (black circles) and for randomised impact careers (grey circles) for six different disciplines. The lack of differences between the two curves in each panel indicates that impact is random within a scientist's sequence of publication for diverse disciplines.

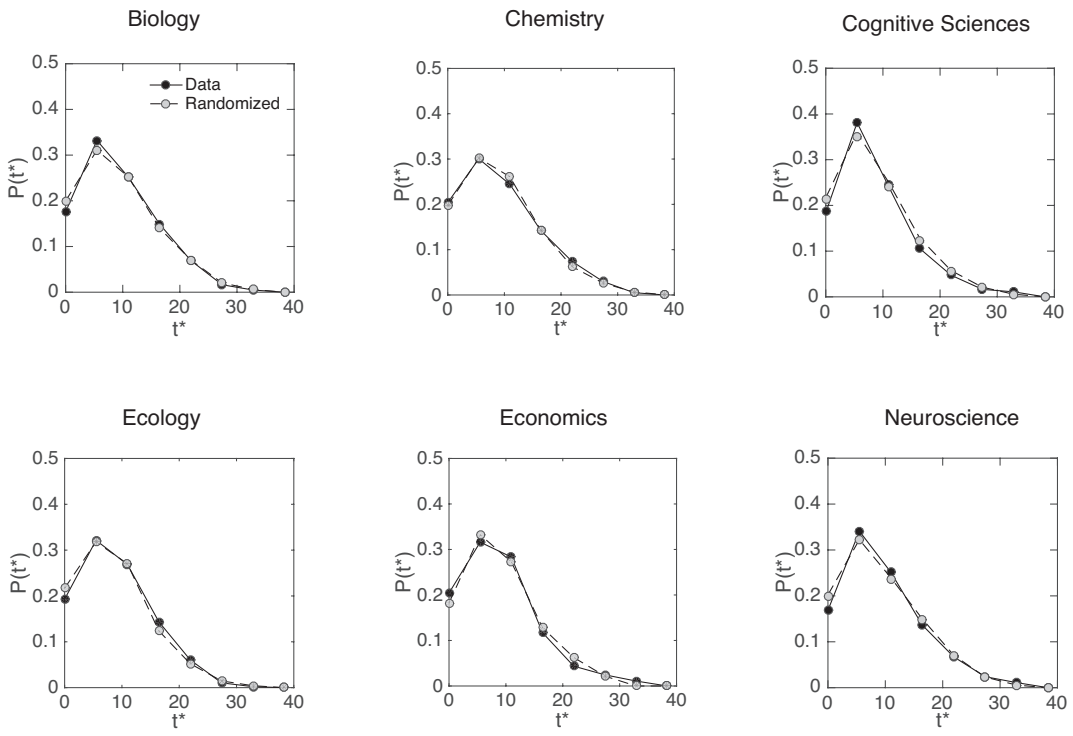

Figure S23: **Timing of the highest impact paper in the six different disciplines of the WoS dataset**. Distribution of the publication time t^* of the highest impact paper for scientists' careers (black circles) and for randomised impact careers (grey circles) for six different disciplines. The lack of differences between the two curves in each panel indicates that impact is random within a scientist's sequence of publication for diverse disciplines.

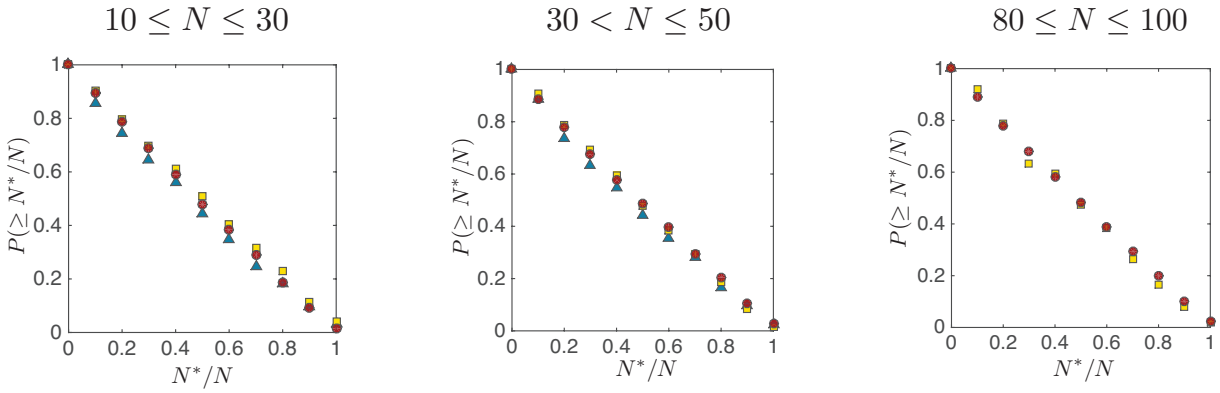

Figure S24: **Control on N for the cumulative distribution $P(\geq N^*/N)$.** Cumulative distribution $P(\geq N^*/N)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career, varying between $1/N$ and 1. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of papers published by a scientist. In each plot, we consider only scientists with similar productivity N, as indicated on the legend on top. No blue triangles are displayed on the right panel as there are not enough low impact scientists with productivity $80 \leq N \leq 100$ to calculate a distribution.

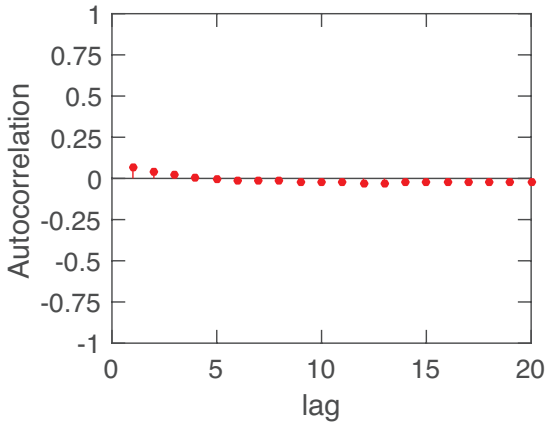

Figure S25: **Impact autocorrelation function in the APS dataset.** We measure the impact autocorrelation function, as described in S3.4, for the APS data. The small values of the autocorrelation (< 0.05 for all lags) further support the random impact rule.

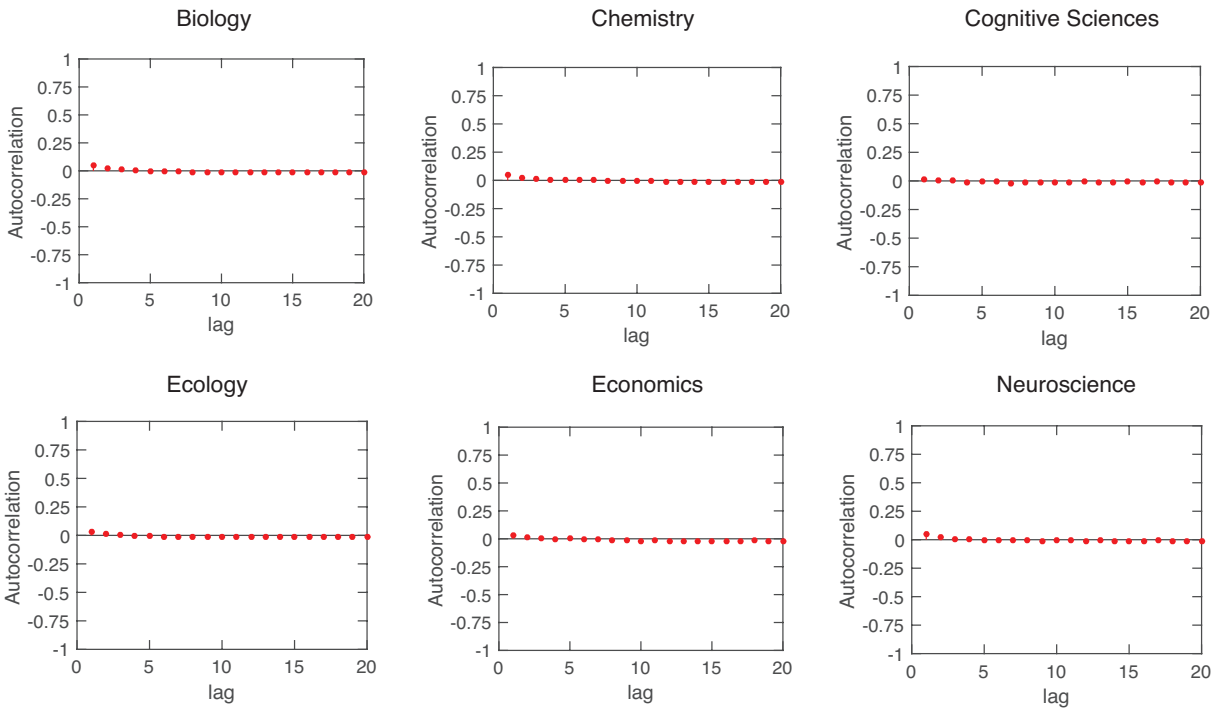

Figure S26: **Impact autocorrelation function in the WoS dataset.** We measure the impact autocorrelation function, as described in S3.4, for the WoS dataset. The small values of the autocorrelation (< 0.05 for all lags) further support the random impact rule.

S8.4 Impact distribution

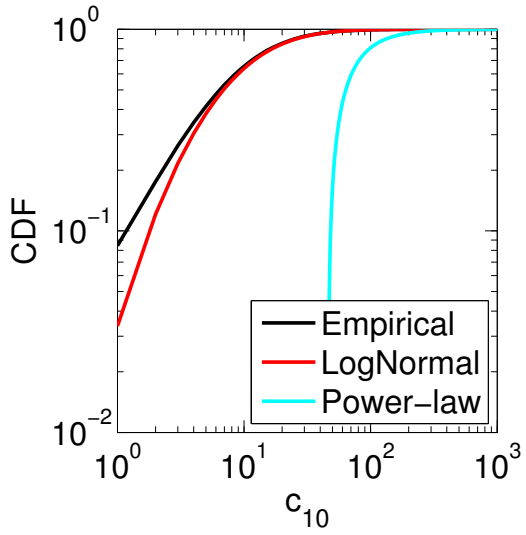

Figure S27: **Lognormal and power-law model for the impact distribution $P(c_{10})$.** We report the cumulative distribution for the empirical data (black line), a lognormal with $\mu = 1.93$ and $\sigma = 1.05$ (red line), and a power-law with exponent $\gamma = 3.13$ and lower bound $c_{10}^{\min} = 49$ (cyan line). The parameters of the lognormal and power-law distribution are estimated by maximizing the likelihood. By using the p-value obtained through a Kolmogorov-Smirnov test we rule out the power-law model in favor of the lognormal model.

S8.5 Goodness of the Q-model and stability of the Q-parameter

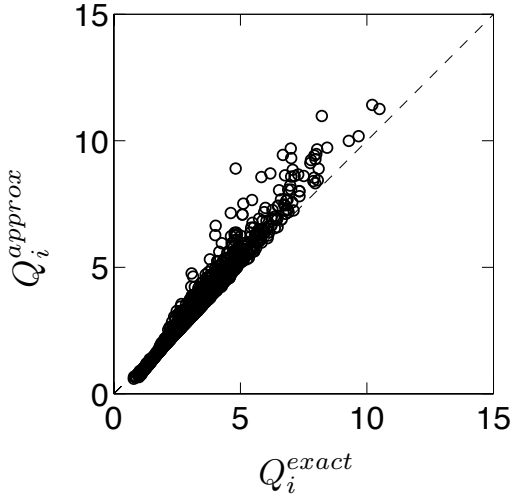

Figure S28: **Approximated vs exact Q-parameter.** Each point of the scatter plot corresponds to a scientist, Q_i^{exact} being the Q-parameter estimated with Eq. (S28), while Q_i^{approx} is estimated with Eq. S30. The pearson correlation coefficient between the two values, $\rho = 0.99$, indicates the almost perfect agreement between the two quantities.

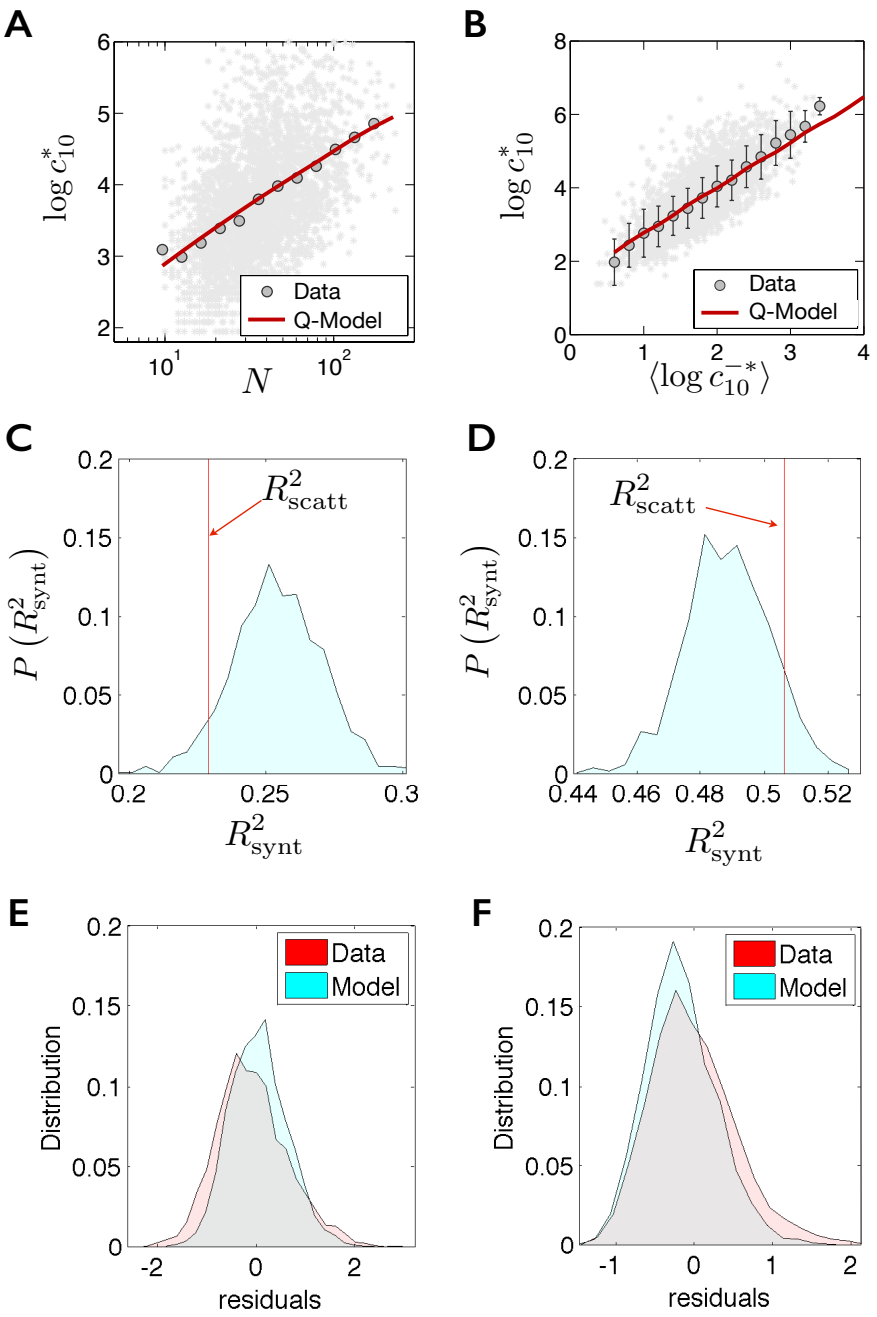

Figure S29: **Goodness of the Q-model to fit the data.** We assess the goodness of the Q-model through the predictions of the scaling of $\langle c_{10}^* \rangle$ **(A)** with productivity N and **(B)** with the average impact excluding the highest impact paper $\langle c_{10}^{-*} \rangle$. **(C)** Distribution of coefficient of discrimination R^2_{synt} between the synthetic scattered data and the Q-model prediction of $\langle c_{10}^* \rangle (N)$. The coefficient of discrimination between the original scattered data R^2_{scatt} and the Q-model falls within the distribution ($p_{value} = 0.23$). **(D)** Same as in (C) but for the prediction $\langle c_{10}^* \rangle \left(\langle c_{10}^{-*} \rangle \right)$. **(E)** Distribution of residuals of the scattered data from the Q-model prediction $\langle c_{10}^* \rangle (N)$. The red area corresponds to the residuals of the data from the prediction, while the blue area corresponds to the residuals of synthetic data. A Mann-Whitney U test rejects the hypothesis that the two sets of residuals belong to the same distribution. However, the Kullback-Leibler distance between the two distributions (2.1% of the maximum distance possible) indicates that very little information is lost when the original data are approximated by the synthetic data (*90*). **(F)** Same as in (E) but for the prediction $\langle c_{10}^* \rangle \left(\langle c_{10}^{-*} \rangle \right)$. In this case, the Kullback-Leibler distance between the two distributions is 1.9% of the maximum distance possible.

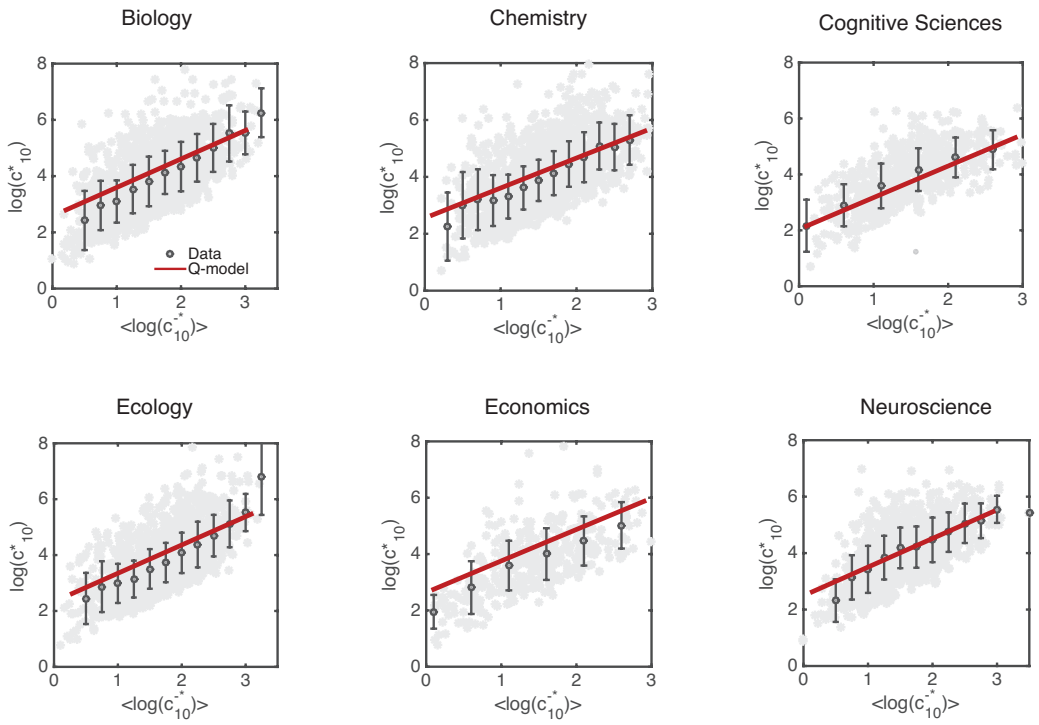

Figure S30: **Prediction of the Q-model for $\log c_{10}^{*}$ vs $\langle \log c_{10}^{-*} \rangle$.** Each grey point in the scatter plot corresponds to a scientist, where $\langle \log c_{10}^{-*} \rangle$ is the average logarithm of his/her paper impact, excluding the most cited paper c_{10}^{*}. We report in red the analytical prediction of the Q-model.

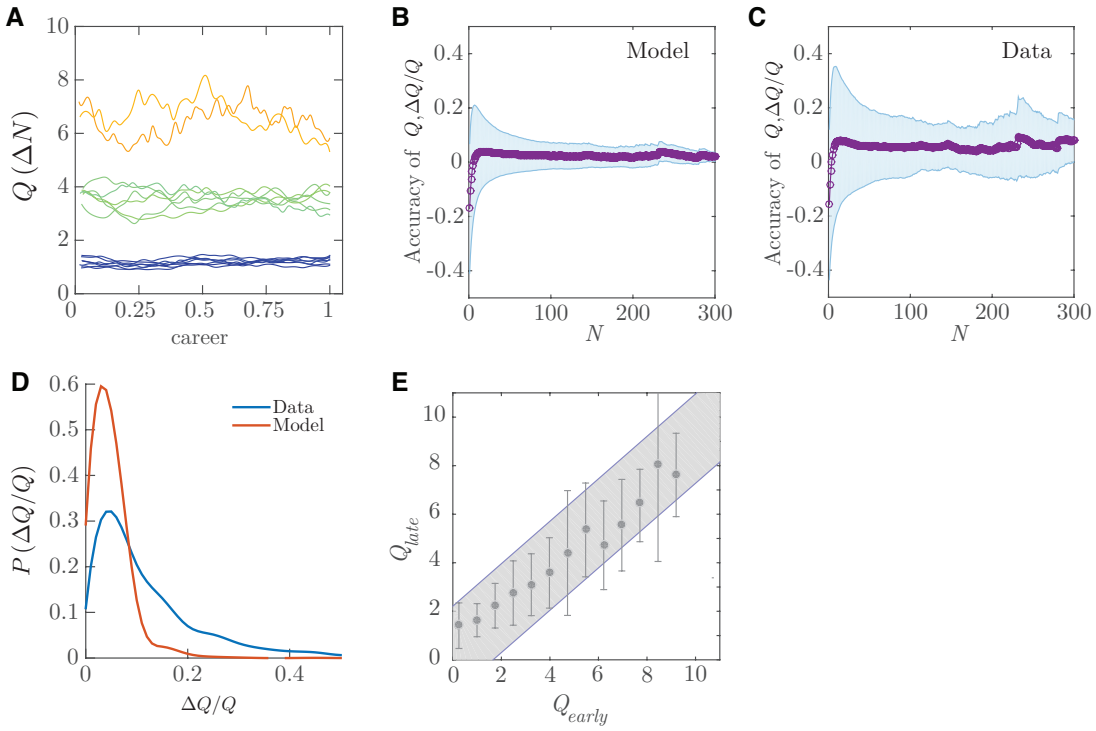

Figure S31: **Accuracy and stability of the Q-parameter**. **(A)** Time variation of the Q-parameter during careers. For scientists with at least 100 papers and $Q \simeq 1.2$, $Q \simeq 3.8$, $Q \simeq 6.5$ and $Q \simeq 9.5$, we report the Q-parameter measured in a moving window of $\Delta N = 30$ papers, $Q\,(\Delta N)$. The fluctuations are mainly due to finite number of papers in the moving window, their magnitude being comparable to the synthetic careers generated with a constant Q parameter. Despite the fluctuations, the Q parameter of scientists in different tiers does not overlap. **(B)** The accuracy on the measurement of the Q-parameter depends on the number of papers available for each scientist. The intrinsic stochastic fluctuations in the estimation of the Q-parameter, ΔQ are predicted by the Fisher Information (*91*). Here we show the prediction for the APS data. **(C)** In the data, the accuracy on Q is measured by the relative error $\frac{\Delta Q}{Q} = \frac{Q(N)-Q}{Q}$, where $Q(N)$ is the Q-parameter estimated with the first N paper published by the scientist, while Q is the value estimated for the entire career. At the early stage of a career, when a scientist has published only a limited number of papers N, the error in the measurement of Q is higher, and decreases with N. **(D)** Fluctuations of the Q-parameter. For each scientist, we study the standard deviation σ of $\frac{Q(\Delta N)}{Q}$, in both data and synthetic careers with constant Q ($\Delta N = 5$). For 74.7% of the scientists, the fluctuations are comparable to those of the model. For the remaining 25.3%, the standard deviation is slightly higher than the one predicted by the model. **(E)** Comparison between early and late Q-parameter. We compare the Q parameter at early (Q_{early}) and late (Q_{late}) career stage of 823 scientists with at least 50 papers. We measured the two values of the parameters using only the first and second half of published papers, respectively. We perform these measurements on the real-data (circles) and on randomized careers, where the order of papers is shuffled (grey shaded areas). For the large majority of careers, 95.1%, the changes in early and stages careers fall within the fluctuations predicted by the null model with randomized paper order, indicating that the Q-parameter is stable throughout a career. The observed fluctuations are explained by the finite number of papers in a scientist's career.

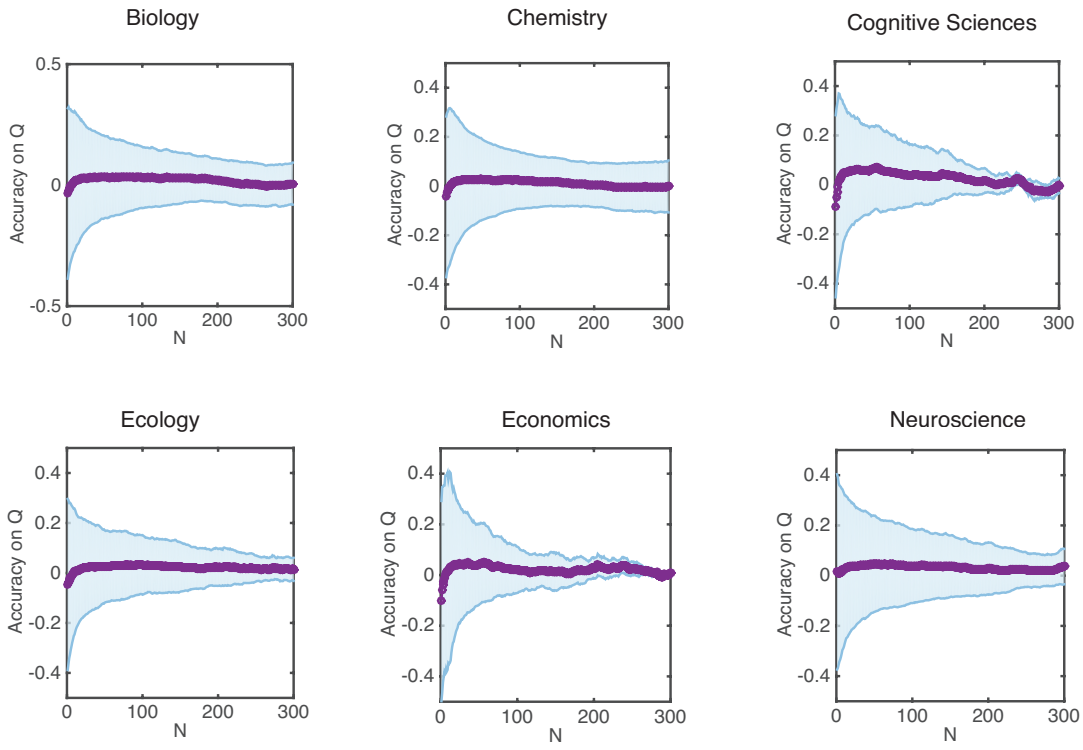

Figure S32: **Relative error on the Q-parameter as a function of the number of papers** N. The accuracy in the measurement of the Q-parameter depends on the number of data points, that is the number of papers, available for each scientist. At the early stage of a career, when a scientist has published only a limited number of papers, the error in the measurement of Q is higher. Here we show the accuracy on Q, measured by the relative error $\frac{Q(N)-\bar{Q}}{\bar{Q}}$, where $Q(N)$ is the Q-parameter estimated with the first N paper published by the scientist, while $\bar{Q}$ is the value estimated for the entire career.

S8.6 Predictive power of the Q-parameter

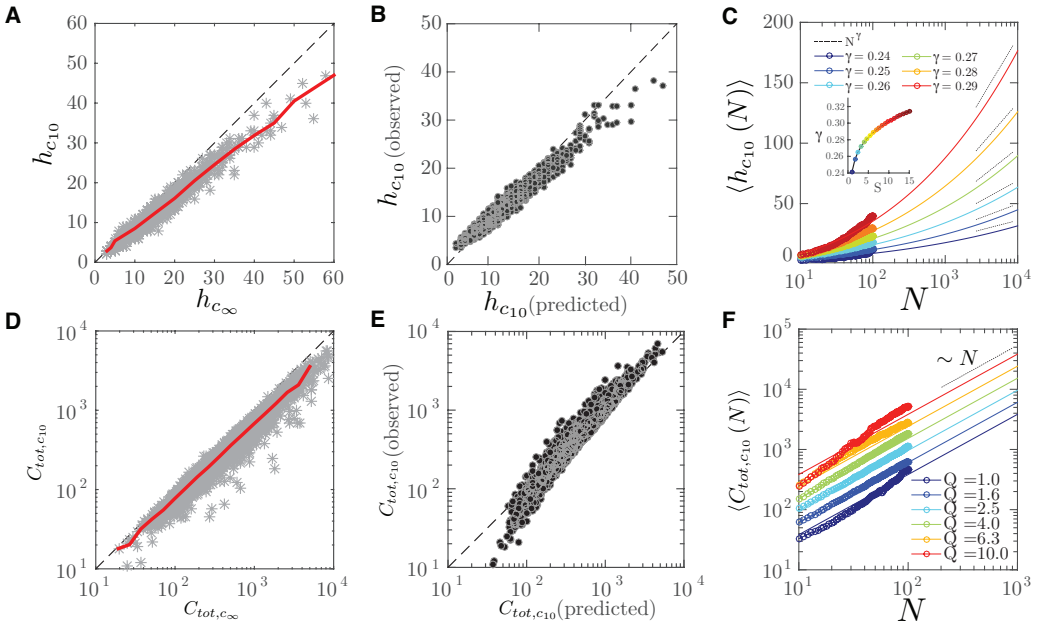

Figure S33: **Relation between impact indicators and the Q-model. (A)** Comparison between the h-index variants $h_{c_{10}}$ and h_{c_∞}. For each scientist, we measure $h_{c_{10}}$, that is the h-index based on the citations a paper has after 10 years, and h_{c_∞}, that is the h-index based on all the citations ever received by a paper. The large majority of papers gathers most of their citations within two or three years after publication *(6, 45)*,yielding $c_{10} \simeq c_\infty$, meaning that most papers contribute in the same way to the calculation of the two variants of the h-index, $h_{c_{10}}$ and h_{c_∞}. However, a small fraction of papers continues to be consistently cited even after 10 years from publication, causing $h_{c_{10}} < h_{c_\infty}$. Being the correlation between the two h-index variants very high (0.98), the differences between $h_{c_{10}}$ and h_{c_∞} can be easily corrected. **(B)** Eq. (S38) allows to predict $h_{c_{10}}$ of each scientist, based on her Q-parameter and productivity N, in excellent agreement with the data (correlation $\rho = 0.98$). **(C)** Observed vs predicted growth of the h-index for scientists with different Q. The plot documents the agreement between the analytically predicted h-index (Eq. S38, continous line) and the observed value $\langle h_{c_{10}}(N)\rangle$, obtained by averaging the h-index for scientists with the same Q (circles). For large N, the h-index scales as N^γ, with γ depending on$Q$, with high Q having a slightly faster scaling (inset). This scaling behavior shows that for scientists with same Q parameter, the h-index is a proxy of productivity N. **(D)** Comparison between $C_{tot,c_{10}}$ and C_{tot,c_∞}. For each scientist, we measure $C_{tot,c_{10}}$, the sum of all the citations his/her papers gather after 10 years, and $C_{tot}(c_\infty)$, considering all the citations ever received by a scientist's papers. As for the the two variants of the h-index in (B), we can have that $C_{tot,c_{10}} < C_{tot,c_\infty}$, since for a small fraction of papers $c_{10} < c_\infty$. However, the high correlation ($rho = 0.94$) indicates that the differences between $C_{tot,c_{10}}$ and C_{tot,c_∞} can be easily corrected. **(E)** Eq. (S39) allows to predict $C_{tot,c_{10}}$ for each scientist, based on her Q-parameter and productivity N, which is in excellent agreement with the data (correlation $\rho = 0.97$). **(F)** Observed vs predicted growth of the total number of citations, $C_{tot,c_{10}}(N)$ for scientists with different Q. The plot documents the agreement between the analytically predicted $C_{tot,c_{10}}$ (Eq. S39, continous line) and the observed value $\langle C_{tot,c_{10}}(N)\rangle$, obtained by averaging the total number of citations for scientists with the same Q (circles). The total number of citation grows linearly with N. For scientists with the same productivity N, the different total number of citations is explained by different values of the Q parameter.

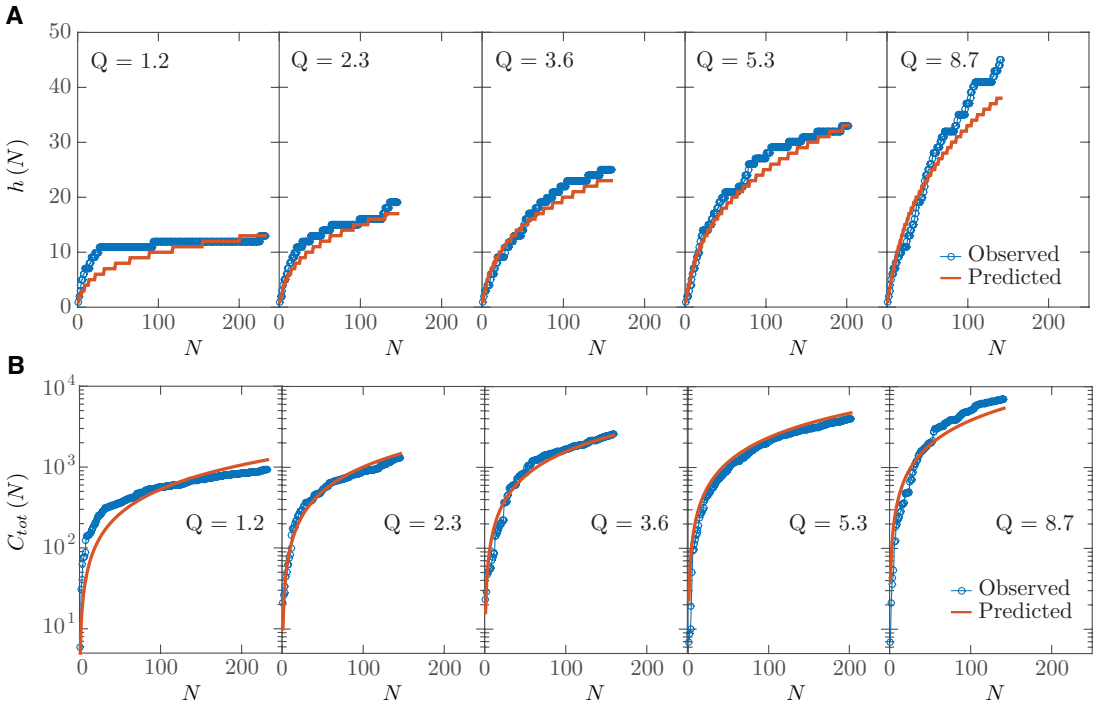

Figure S34: Q**-model prediction of impact indicator dynamic. (A)** The growth of the h-index for five scientists with at least 140 papers and different Q as a function of the productivity N (blu circles), compared it with the prediction of Eq. (S38) (red line). The first and fourth panel are the same as in Fig. 4E. **(B)** For the five scientists shown in (A), we measure the cumulative number of citations $C_{tot}\,(N)$ as a function of N and compare it with the prediction of Eq. (S39).

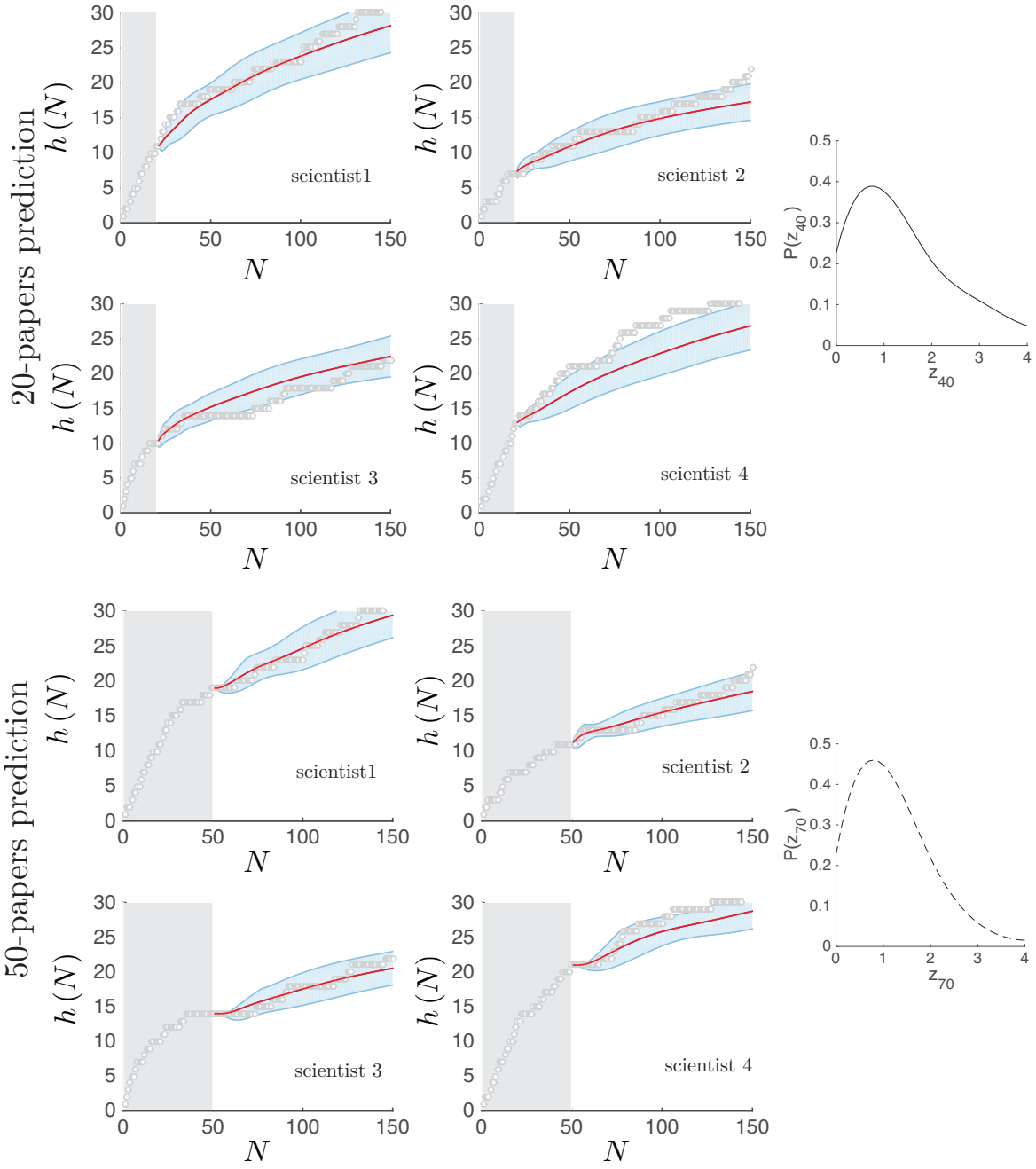

Figure S35: **Prediction of the h-index using early career information.** For 4 scientists with 200 papers in their career, we measure $h(c_{10})$, that is the h-index based on the citations papers gather after 10 years. We measure also the Q- parameter using only the first 20 (top) and 50 (bottom) papers published (grey area) and use this value in Eqs. (S44)-(S45)) to determine $h_{c_{10}}$. For scientists 1-3 the prediction based on 20 papers encloses the entire observed h-index evolution. For scientist 4, a more accurate prediction of the Q parameter, based on 50 papers, is necessary to have the observed data falling within the prediction envelope. We also report in the panels on the right the distribution of the z-score of the observed h-index vs the predicted at $N = 40$ (top) and $N = 70$ (bottom). The distribution $P(z)$ indicate that for most of the scientists (74% for $N_0 = 20$ and $N = 40$, and 81% for $N_0 = 50$ and $N = 70$) $z \leq 2$, that is the observed data falls within the uncertainty envelop of the prediction. The close agreement between observed and prediction $h(N)$ shows that the Q-parameter can be estimated early in a scientist's career and can be used for accurate predictions of future impact.

S8.7 Robustness of the core results for different dataset selections

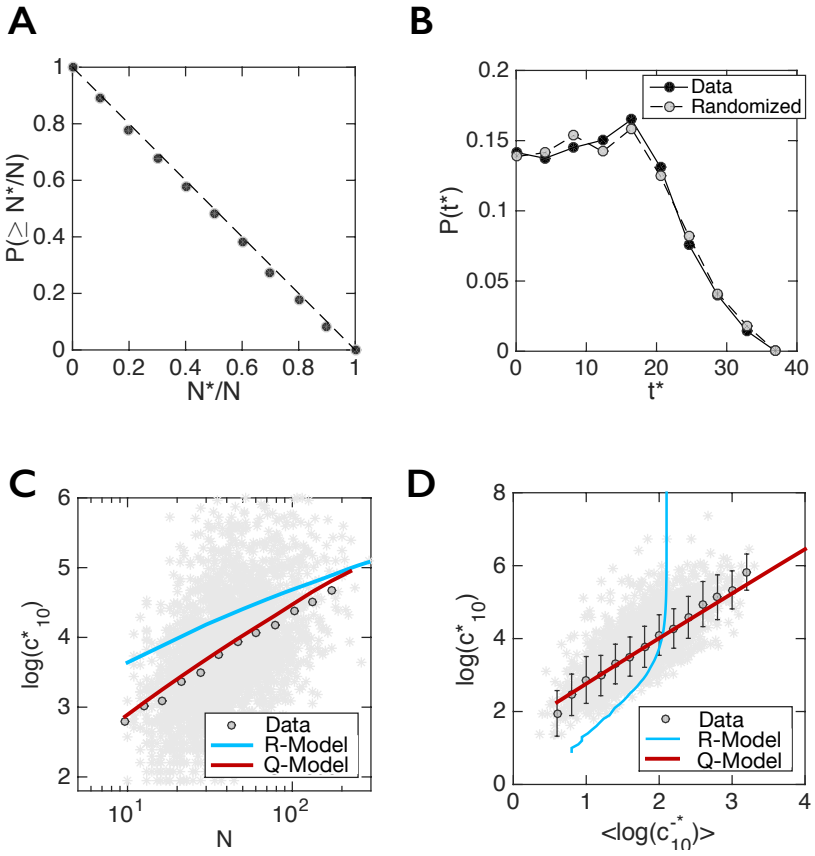

Figure S36: **Randomness of the highest impact paper and Q-model predictions for the dataset with no review papers.** We consider the 2,887 careers studied in the main article, and remove all the review papers published in Review of Modern Physics and repeat the following measurements: **(A)** Cumulative distribution $P\left(\geq N^*/N\right)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of papers published by a scientist; **(B)** Distribution of the publication time t^* of the highest impact paper c_{10}^* for scientists' careers (black circles) and for randomized impact careers (grey circles). **(C)** Citations of the highest impact paper, c_{10}^*, vs the number of publications N during a scientist's career. The circles are the logarithmic binning of the scattered data, the cyan curve represents the prediction of the R-model, assuming that the impact of each paper is extracted randomly from the distribution $P\left(c_{10}\right)$ and the red curve corresponds to the analytical prediction of the Q-model; **(D)** $\log c_{10}^*$ vs $\langle \log c_{10}^{-*} \rangle$, where $\langle \log c_{10}^{-*} \rangle$ is the average logarithm of the impact of a scientist's papers excluding the highest impact work c_{10}^*. We report in cyan the R-model prediction and in red the analytical prediction of the Q-model. All the measurements show that our findings remain unchanged when removing the review papers, indicating that our results are not sensitive to the different impact distribution of review papers.

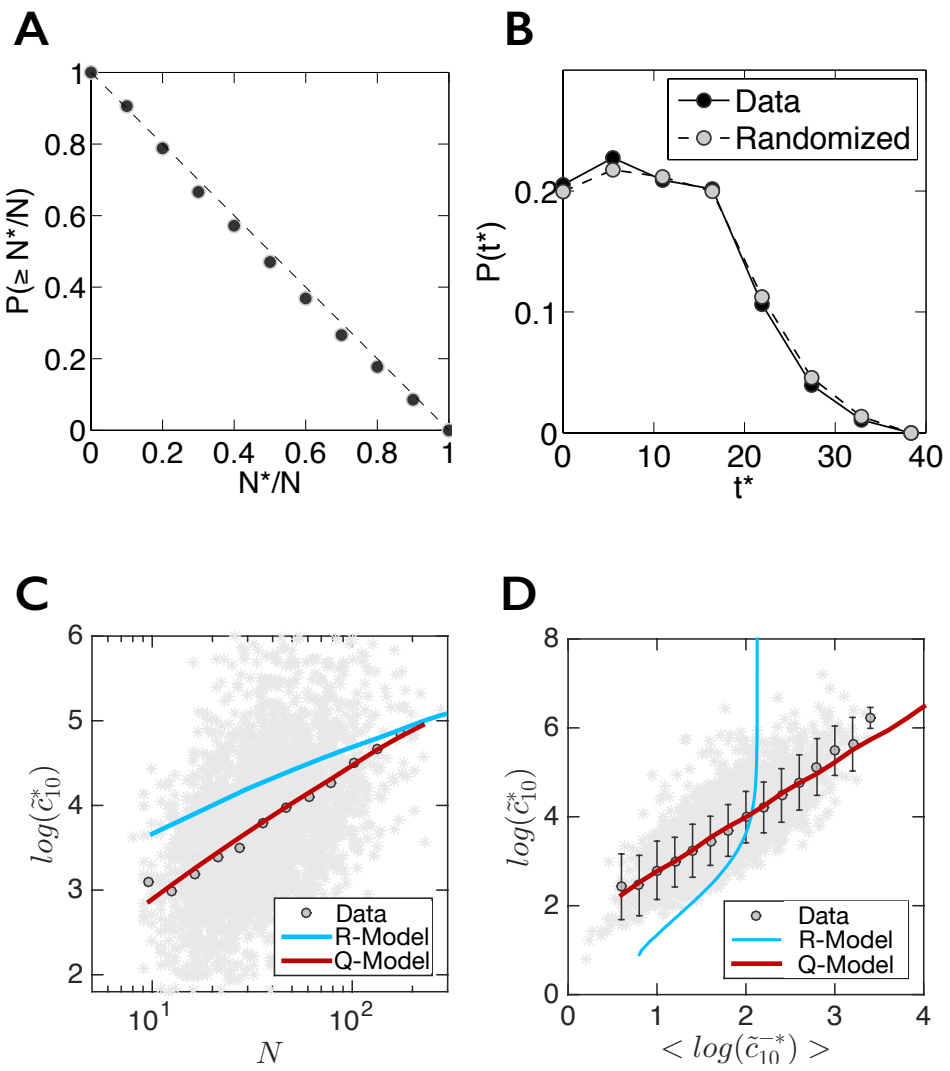

Figure S37: **Randomness of the highest impact paper and Q-model predictions using the time rescaled impact $\tilde{c}_{10}$.** We use the rescaled impact $\tilde{c}_{10}$ defined in Eq. (S1) and repeat the following measurements: **(A)** Cumulative distribution $P\left(\geq N^*/N\right)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of papers published by a scientist; **(B)** Distribution of the publication time t^* of the highest impact paper $\tilde{c}^*_{10}$ for scientists' careers (black circles) and for randomized impact careers (grey circles). **(C)** Citations of the highest rescaled impact paper, $\tilde{c}^*_{10}$, vs the number of publications N during a scientist's career. The circles are the logarithmic binning of the scattered data, the cyan curve represents the prediction of the R-model, assuming that the impact of each paper is extracted randomly from the distribution $P\left(c_{10}\right)$ and the red curve corresponds to the analytical prediction Eq. (S35) of the Q-model; **(D)** $\log c^*_{10}$ vs $\langle \log c^{-*}_{10} \rangle$, where $\langle \log c^{-*}_{10} \rangle$ is the average logarithm of the impact of a scientist's papers excluding the highest impact work c^*_{10}. We report in cyan the R-model prediction and in red the analytical prediction Eq. (S35) of the Q-model. All the measurements show that our findings remain unchanged when using the time rescaled impact $\tilde{c}_{10}$, indicating that they are robust to the changes of the average number of citations per paper over time.

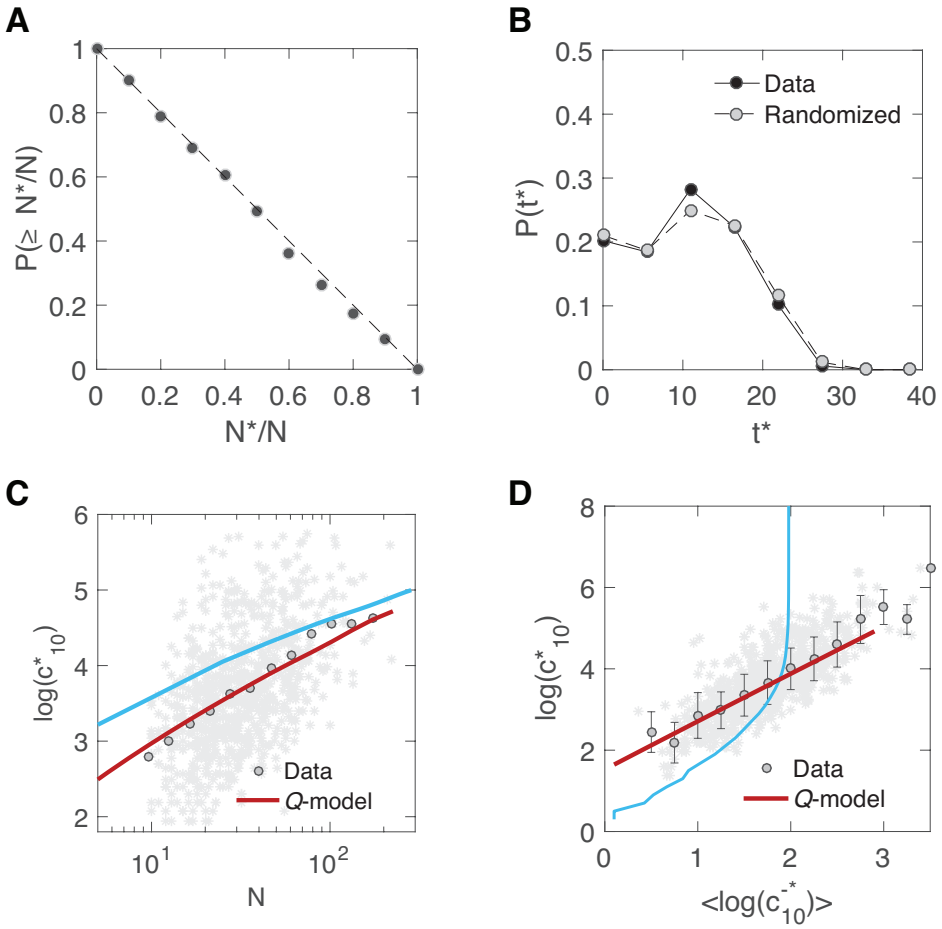

Figure S38: **Randomness of the highest impact paper and Q-model predictions using the PRB dataset.** We use the dataset described in S38 and repeat for the scientists studied in the main text the following measurements: **(A)** Cumulative distribution $P\left(\geq N^*/N\right)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of papers published by a scientist; **(B)** Distribution of the publication time t^* of the highest impact paper c_{10}^* for scientists' careers (black circles) and for randomized impact careers (grey circles). **(C)** Citations of the highest rescaled impact paper, c_{10}^*, vs the number of publications N during a scientist's career. The circles are the logarithmic binning of the scattered data, the cyan curve represents the prediction of the R-model, assuming that the impact of each paper is extracted randomly from the distribution $P\left(c_{10}\right)$ and the red curve corresponds to the analytical prediction Eq. (S35) of the Q-model; **(D)** $\log c_{10}^*$ vs $\langle \log c_{10}^{-*} \rangle$, where $\langle \log c_{10}^{-*} \rangle$ is the average logarithm of the impact of a scientist's papers excluding the highest impact work c_{10}^*. We report in cyan the R-model prediction and in red the analytical prediction Eq. (S35) of the Q-model. All the measurements show that our findings remain unchanged when considering only publications in one area of physics.

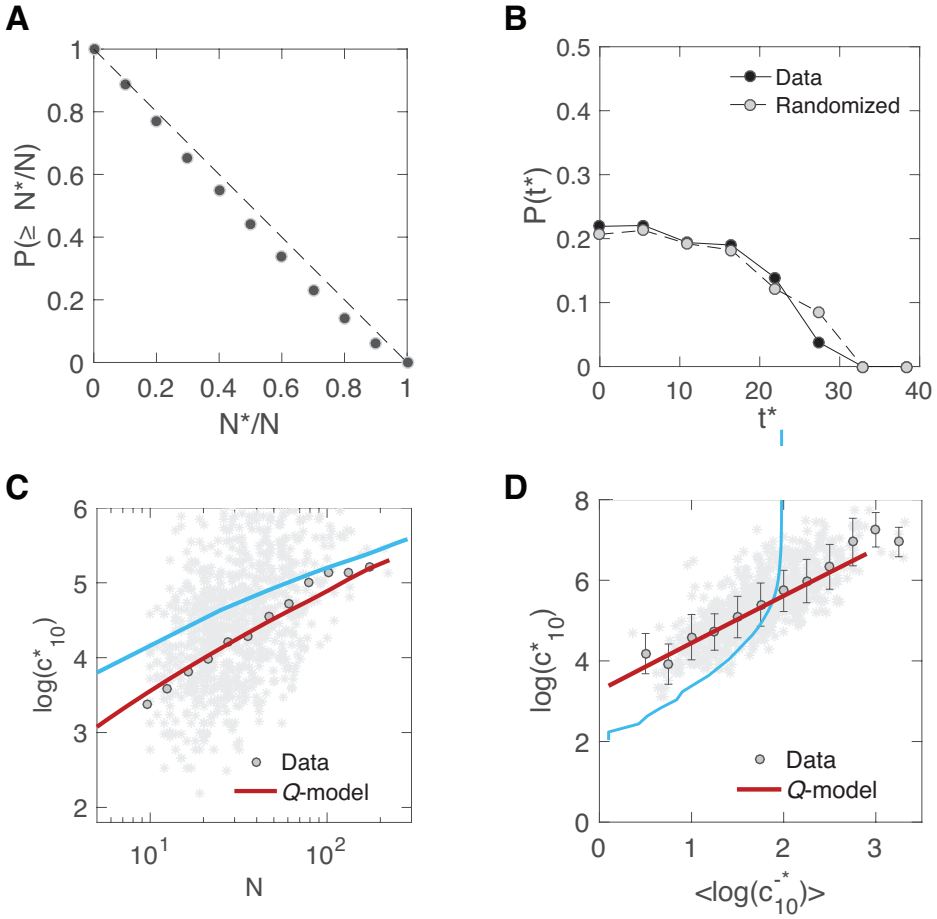

Figure S39: **Randomness of the highest impact paper and Q-model predictions using data from WoS for the APS journal.** We use the dataset described in S1.5 and repeat for the scientists studied in the main text the following measurements: **(A)** Cumulative distribution $P\left(\geq N^*/N\right)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career. **(B)** Distribution of the publication time t^* of the highest impact paper c_{10}^* for scientists' careers (black circles) and for randomized impact careers (grey circles). **(C)** Citations of the highest rescaled impact paper, c_{10}^*, vs the number of publications N during a scientist's career. The circles are the logarithmic binning of the scattered data, the cyan curve represents the prediction of the R-model, assuming that the impact of each paper is extracted randomly from the distribution $P\left(c_{10}\right)$ and the red curve corresponds to the analytical prediction Eq. (S35) of the Q-model; **(D)** $\log c_{10}^*$ vs $\langle \log c_{10}^{-*} \rangle$, where $\langle \log c_{10}^{-*} \rangle$ is the average logarithm of the impact of a scientist's papers excluding the highest impact work c_{10}^*. We report in cyan the R-model prediction and in red the analytical prediction Eq. (S35) of the Q-model. All the measurements show that our findings remain unchanged when considering APS publications in WoS.

S8.8 Robustness of the core results when controlling for coauthorship effects

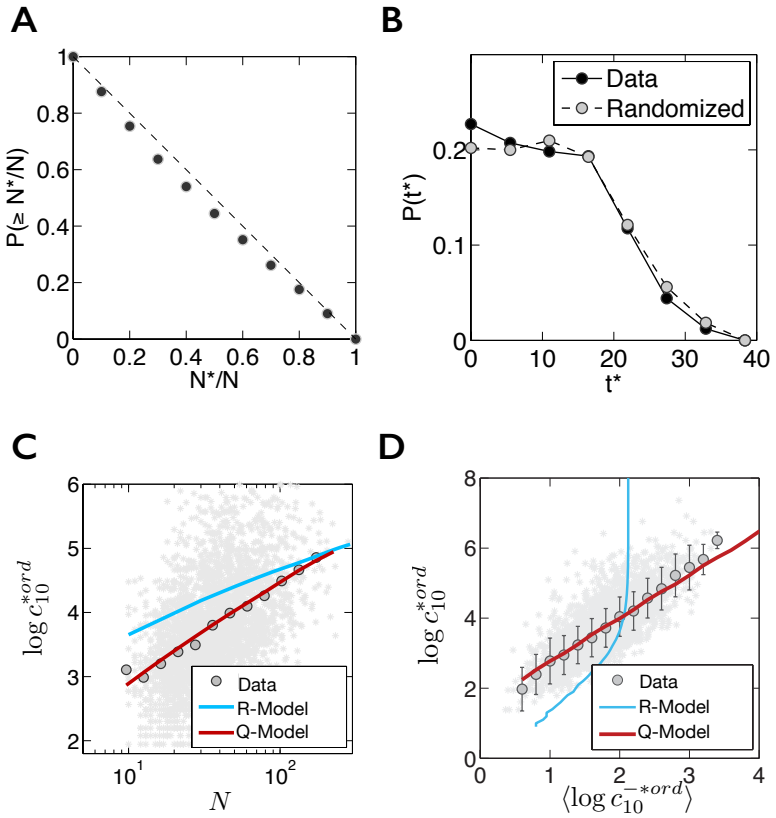

Figure S40: **Empirical observations for impact assigned based on the author list order.** Each paper α in a scientist i's career is associated with an impact $c^{ord}_{10,i\alpha} = c_{10,i\alpha} q_i^{\alpha}$, that is a fraction q_i^{α} of citations accumulated by α in 10 years, $c_{10,ip}$. The fraction q_i^{α} is computed based on author i's rank in the authorlist of α *(36)*. **(A)** Cumulative distribution $P(\geq N^*/N)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career, varying between 1/N and 1. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of papers published by a scientist. **(B)** Distribution of the publication time t^* of the highest impact paper $c_{10}^{ord,*}$ for scientists' careers (black circles) and for randomized impact careers (grey circles). The lack of differences between the two curves confirms the random impact rule. **(C)** Citations of the highest impact paper, $c_{10}^{*,ord}$, vs the number of publications N during a scientist's career. The circles are the logarithmic binning of the scattered data, the cyan curve represents the prediction of the R-model, assuming that the impact of each paper is extracted randomly from the distribution $P(c_{10})$ and the red curve corresponds to the analytical prediction of the Q-model with corrected parameters $\mu = (0.80, 0.5, 3.4)$ and $\Sigma = \begin{pmatrix} 0.91 & 0 & 0 \\ 0 & 0.12 & 0.18 \\ 0 & 0.18 & 0.33 \end{pmatrix}$. **(D)** $\log c_{10}^{*,ord}$ vs $\langle \log c_{10}^{-*,ord} \rangle$, where $\langle \log c_{10}^{-*,ord} \rangle$ is the average logarithm of the order based impact of a scientist's papers excluding the highest impact work $c_{10}^{*,ord}$. We report the R-model prediction (cyan curve) and the Q-model analytical prediction with the parameters reported in (C) (red curve).

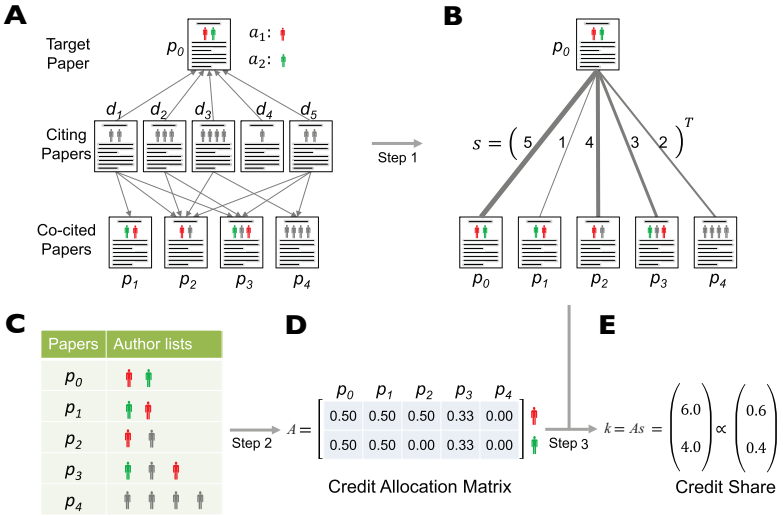

Figure S41: **Illustrating the credit allocation process**. **(A)** The target paper p_0 has two authors, a_1 and a_2, colored in red and green respectively. We also show the citing papers d_k $(1 \leq k \leq 5)$ and the co-cited papers p_j $(0 \leq j \leq 4)$ that were cited by these citing papers together with p_0. **(B)** The p_0-centric co-citation network constructed from (A), where the weights of links denote the co-citation strength s between the co-cited papers and the target paper p_0. **(C)** The author lists of the target paper p_0 and its co-cited papers. **(D)** The credit allocation matrix $\mathbf{A}$ obtained from the author lists of the co-cited papers in (C). The matrix $\mathbf{A}$ provides for each co-cited paper the authors' share. For example, since p_2 has a_1 as one of its two authors but it lacks the author a_2, it votes 0.5 for author a_1 and 0.0 for author a_2. **(E)** With the matrix $\mathbf{A}$ and co-citation strength $\mathbf{s}$, the credit share of the two authors of p_0 is computed according to Eq. (S46) with a normalization. Reprinted figure with permission of the authors from *(22)*.

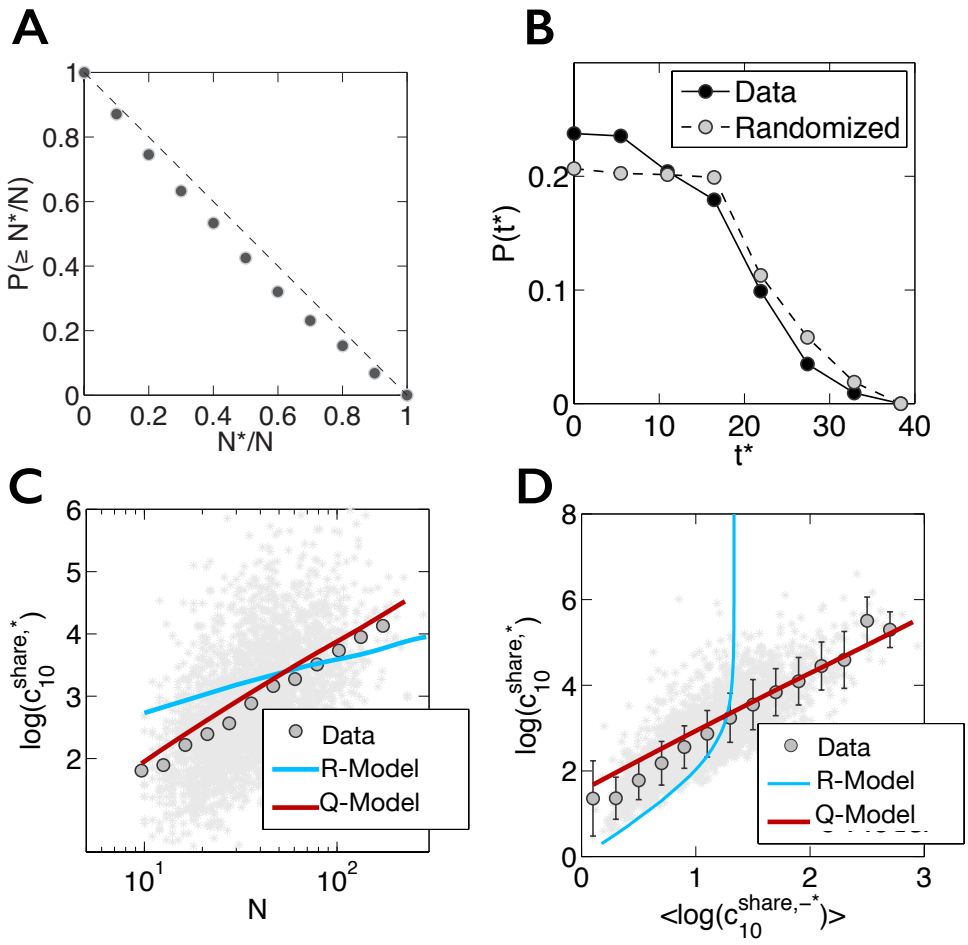

Figure S42: **Empirical observations for impact assigned proportionally to the credit share.** Each paper α in a scientist i's career, is associated with an impact $c_{10,i\alpha}^{share} = c_{10,i\alpha}k_i^\alpha$, that is a fraction k_i^α of $c_{10,i\alpha}$, that is citations 10 years after the publication of α. k_i^α is calculated according to Eq. S46 and then normalized. **(A)** Cumulative distribution $P\left(\geq N^*/N\right)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career, varying between 1/N and 1. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that $N*$ has the same probability to occur anywhere in the sequence of papers published by a scientist. **(B)** Distribution of the publication time t^* of the highest impact paper $c_{10}^{share,*}$ for scientists' careers (black circles) and for randomized impact careers (grey circles). The lack of differences between the two curves confirms the random impact rule. **(C)** citations of the highest rescaled impact paper, $c_{10}^{share,*}$, vs the number of publications N during a scientist's career. The circles are the logarithmic binning of the scattered data, the cyan curve represents the prediction of the R-model, assuming that the impact of each paper is extracted randomly from the distribution $P\left(c_{10}\right)$ and the red curve corresponds to the analytical prediction of the Q-model with parameters $\mu = (0.73, 0.4, 3.4)$ and $\Sigma = \begin{pmatrix} 0.93 & 0 & 0 \\ 0 & 0.15 & 0.14 \\ 0 & 0.14 & 0.32 \end{pmatrix}$. Note that $\sigma_{Q,N}$ is larger than in the original dataset and of the same order as σ_Q^2 **(D)** $\log c_{10}^{share,*}$ vs $\langle \log c_{10}^{share,-*} \rangle$, where $\langle \log c_{10}^{share,-*} \rangle$ is the average logarithm of the impact of a scientist's papers excluding the highest impact work c_{10}^*.We report the R-model prediction (cyan) and the Q-model analytical prediction with the parameters reported in (E) (red).

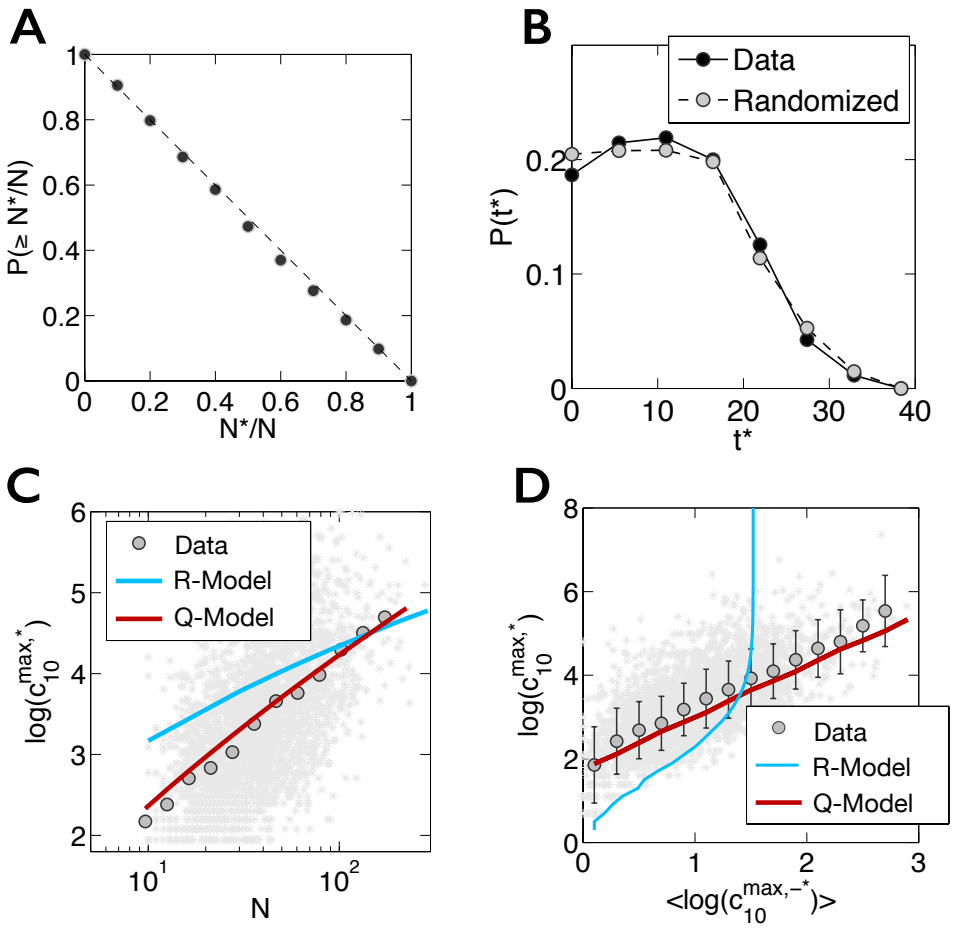

Figure S43: **Empirical observations for impact based on maximum credit share.** Each paper α in a scientist i's career, is associated with an impact $c_{10,i\alpha}^{max}$ which is equal to the original impact $c_{10,i\alpha}$ if i has the maximum credit share k_i^α among all the authors of α, otherwise $c_{10,i\alpha}^{max} = 0$. If two or more authors have the same maximum credit share, they all get for α the original impact $c_{10,\alpha}$. **(A)** Cumulative distribution $P\left(\geq N^*/N\right)$, where N^*/N denotes the order N^* of the highest impact paper in a scientist's career, varying between 1/N and 1. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of papers published by a scientist. **(B)** Distribution of the publication time t^* of the highest impact paper $c_{10}^{max,*}$ for scientists' careers (black circles) and for randomized impact careers (grey circles). The lack of differences between the two curves confirms the random impact rule. **(C)** Citations of the highest impact paper, $c_{10}^{max,*}$, vs the number of publications N during a scientist's career. The circles are the logarithmic binning of the scattered data, the cyan curve represents the prediction of the R-model, assuming that the impact of each paper is extracted randomly from the distribution $P\left(c_{10}\right)$ and the red curve corresponds to the analytical prediction of the Q-model with parameters $\mu = (0.73, 0.15, 3.4)$ and $\Sigma = \begin{pmatrix} 0.93 & 0 & 0 \\ 0 & 0.15 & 0.09 \\ 0 & 0.09 & 0.32 \end{pmatrix}$. **(D)** $\log c_{10}^*$ vs $\langle \log c_{10}^{-*} \rangle$, where $\langle \log c_{10}^{-*} \rangle$ is the average logarithm of the impact of a scientist's papers excluding the highest impact work c_{10}^*. We report the R-model prediction (cyan curve) and the Q-model analytical prediction with the parameters reported in (E) (red curve).

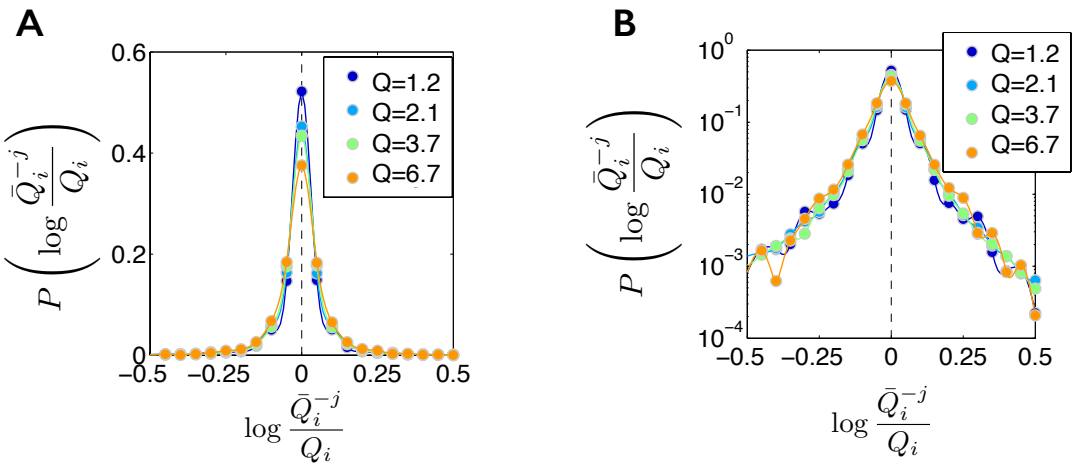

Figure S44: **Q-parameter variations induced by coauthors.** **(A)** Distribution of the Q-parameter variation $P(\log \frac{\bar{Q}_i^{-j}}{Q_i})$ for scientists with the same Q parameter. For each scientist i, we calculate the ratio $\log \frac{\bar{Q}_i^{-j}}{Q_i}$, where $\bar{Q}_i^{-j}$ is computed using the scientist's subset of publications lacking all papers written by co-author j, while Q_i is calculated based on the entire body of publications of i. Notice that $\log \frac{\bar{Q}_i^{-j}}{Q_i} = 0$ indicates no change in Q, while for $\log \frac{\bar{Q}_i^{-j}}{Q_i} > 0$ the parameter increases and for $\log \frac{\bar{Q}_i^{-j}}{Q_i} < 0$ the parameter decreases as a result of removing publications written with coauthor j. Each distribution is computed for scientists with similar Q, reported in the legend. For all Q, the distribution is symmetric and peaked at zero, indicating that there are no systematic changes in the computation of the Q when we remove individual co-authors. **(B)** Same as in (A) but on a lin-log plot.

S8.9 Prediction of independent recognitions

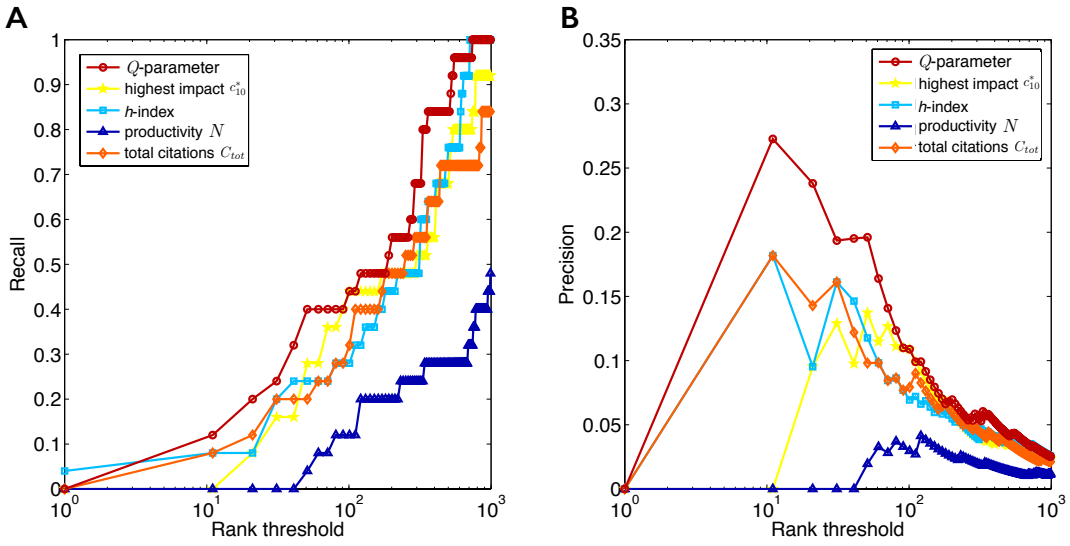

Figure S45: **Recall and precision for the rankings of Nobel laureates**. **(A)** The recall, defined in section S7, for the rankings of Nobel laureates based on Q, c_{10}^*, h-index, N, and C_{tot} (ROC-plot shown in Fig. MS-fig3D of the main text). The plot indicates that the Q parameter has the best recall over all other rankings for all rank thresholds. **(B)** The precision, defined in section S7, for the rankings of Nobel laureates based on Q, c_{10}^*, h-index, N, and C_{tot}. The Q-parameter has the best precision for all rank thresholds.

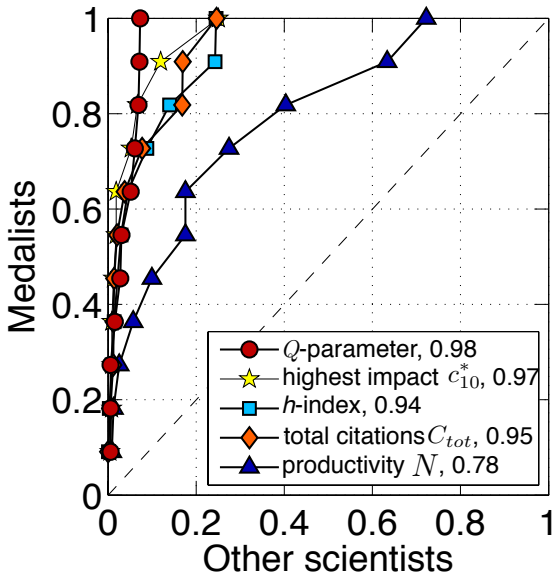

Figure S46: **ROC-plot for Dirac and Boltzmann medalists**. ROC-plot capturing the ranking of scientists based on Q, h-index, c_{10}^*, N, and C_{tot}. Each curve represents the fraction of Dirac and Boltzmann medalists vs the fraction of other scientists for different rank thresholds. Similarly to the case of Nobel Laureates, the ranking based on Q has the best accuracy, equal to 0.98. The accuracy for all rankings are reported in the legend.

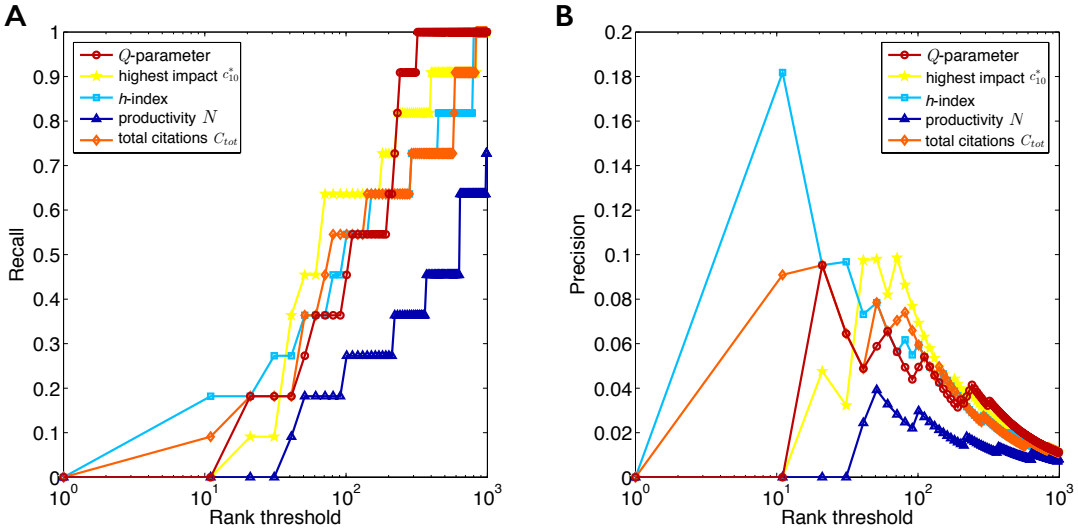

Figure S47: **Recall and precision for the rankings of Dirac and Boltzmann medalists**. **(A)** The recall, defined in section S7, based on Q, c_{10}^*, h-index, N, and C_{tot} (ROC-plot in Fig.S46). The Q-parameter has a low recall up to the rank threshold of 10, after which it increases steeply up to having the best value after a rank threshold of 120. **(B)** The precision, defined in section S7, based on Q, c_{10}^*, h-index, N, and C_{tot}. The Q-parameter has slightly lower precision in respect to other rankings for values of the rank threshold up to ~ 300. After this threshold, Q yields the best precision.

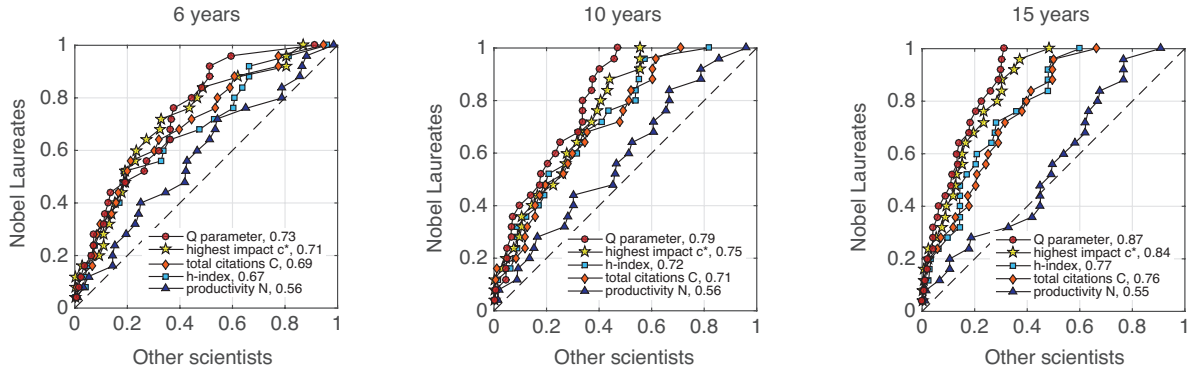

Figure S48: **ROC plots based on early estimation of the Q-parameter.** ROC-plot capturing the ranking of scientists based on Q, C_{tot}, h-index, c^*_{10}, and N estimated in the first 6 (left), 10 (center) and 15 (right) years of career. Each curve represents the fraction of Nobel laureates vs the fraction of other scientists for a given rank threshold. The diagonal (no-discrimination line) corresponds to random ranking; the area under each curve provides our accuracy to rank high Nobel laureates. The rankings accuracy is reported in the legend, 1 being the maximum. At each career stage the Q-parameter is the best predictor of becoming a Nobel Laureate.

S8.10 Early impact vs productivity

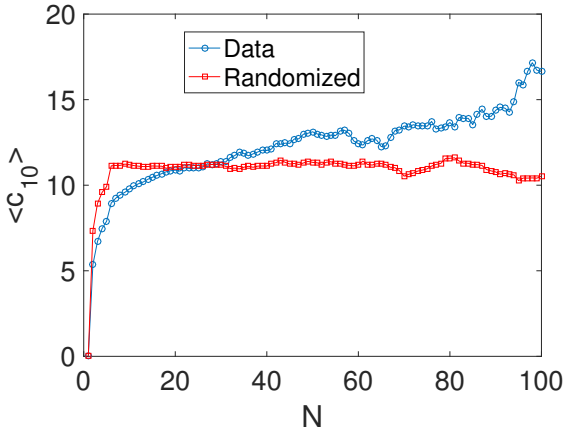

Figure S49: **Average early impact vs career longevity.** We measure the average impact $\langle c_{10} \rangle$ for the first 10 papers published by a scientist (hence not considering scientists with less than 10 papers) as a function of the career longevity, quantified by the total number of papers published N, and compared it with that of randomized careers. We find that there average early impact has a small effect on the overall career longevity. Indeed doubling productivity from 30 to 60 papers is associated to an increase of 3 citations for $\langle c_{10} \rangle$. This indicates that there is a effect, although not pronounced, between early impact and probability to stay long in the scientific enterprise.

References

1. I. Fuyuno, D. Cyranoski, Cash for papers: Putting a premium on publication. *Nature* **441**, 792 (2006). doi:10.1038/441792b; Medline

2. J. A. Evans, J. Reimer, Open access and global participation in science. *Science* **323**, 1025 (2009). doi:10.1126/science.1154562; Medline

3. P. Azoulay, Research efficiency: Turn the scientific method on ourselves. *Nature* **484**, 31–32 (2012). doi:10.1038/484031a; Medline

4. B. Owens, Research assessments: Judgement day. *Nature* **502**, 288–290 (2013). doi:10.1038/502288a; Medline

5. O. Penner, R. K. Pan, A. M. Petersen, K. Kaski, S. Fortunato, On the predictability of future impact in science. *Sci. Rep.* **3**, 3052 (2013). doi:10.1038/srep03052; Medline

6. D. Wang, C. Song, A.-L. Barabási, Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013). doi:10.1126/science.1237825; Medline

7. J. A. Evans, Future science. *Science* **342**, 44–45 (2013). doi:10.1126/science.1245218; Medline

8. S. Lehmann, A. D. Jackson, B. E. Lautrup, Measures of measures. *Nature* **444**, 1003–1004 (2006). doi:10.1038/4441003a; Medline

9. S. Lehmann, A. D. Jackson, B. E. Lautrup, A quantitative analysis of indicators of scientific performance. *Scientometrics* **76**, 369–390 (2008). doi:10.1007/s11192-007-1868-8

10. F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, Diffusion of scientific credits and the ranking of scientists. *Phys. Rev. E* **80**, 056103 (2009). doi:10.1103/PhysRevE.80.056103; Medline

11. D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, I. Rafols, Bibliometrics: The Leiden Manifesto for research metrics. *Nature* **520**, 429–431 (2015). doi:10.1038/520429a; Medline

12. J. E. Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16569–16572 (2005). doi: 10.1073/pnas.0507655102; Medline

13. S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, F. Herrera, h-Index: A review focused in its variants, computation and standardization for different scientific fields. *J. Informet.* **3**, 273–289 (2009). doi:10.1016/j.joi.2009.04.001

14. L. Bornmann, R. Mutz, S. E. Hug, H.-D. Daniel, A multilevel meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants. *J. Informet.* **5**, 346–359 (2011). doi:10.1016/j.joi.2011.01.006

15. D. E. Acuna, S. Allesina, K. P. Kording, Future impact: Predicting scientific success. *Nature* **489**, 201–202 (2012). doi:10.1038/489201a; Medline

16. B. F. Jones, B. A. Weinberg, Age dynamics in scientific creativity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18910–18914 (2011). doi:10.1073/pnas.1102895108; Medline

17. P. Azoulay, J. S. Graff Zivin, G. Manso, Incentives and creativity: Evidence from the academic life sciences. *RAND J. Econ.* **42**, 527–554 (2011). doi:10.1111/rand.2011.42.issue-3

18. A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli, Persistence and uncertainty in the academic career. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5213–5218 (2012). doi:10.1073/pnas.1121429109

19. C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H. E. Stanley, Long-range correlations in nucleotide sequences. *Nature* **356**, 168–170 (1992). doi:10.1038/356168a0; Medline

20. D. K. Simonton, Career landmarks in science: Individual differences and interdisciplinary contrasts. *Dev. Psychol.* **27**, 119–130 (1991). doi:10.1037/0012-1649.27.1.119

21. D. K. Simonton, Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychol. Rev.* **104**, 66–89 (1997). doi:10.1037/0033-295X.104.1.66

22. H.-W. Shen, A.-L. Barabási, Collective credit allocation in science. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12325–12330 (2014). doi:10.1073/pnas.1401992111; Medline

23. D. de Solla Price, *Little Science, Big Science… and Beyond* (Columbia University, 1963).

24. F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: Toward an objective measure of scientific impact. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17268–17272 (2008). doi:10.1073/pnas.0806977105; Medline

25. M. J. Stringer, M. Sales-Pardo, L. A. Nunes Amaral, Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *J. Am. Soc. Inf. Sci. Technol.* **61**, 1377–1385 (2010). doi:10.1002/asi.v61:7; Medline

26. E. J. Gumbel, *Statistics of Extremes* (Dover Publications, 1958).

27. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer-Verlag, 2001).

28. D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, H. A. Makse, Scaling laws of human interaction activity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12640–12645 (2009). doi:10.1073/pnas.0902667106; Medline

29. D. Garlaschelli, G. Caldarelli, L. Pietronero, Universal scaling relations in food webs. *Nature* **423**, 165–168 (2003). doi:10.1038/nature01604; Medline

30. T. Vicsek, A question of scale. *Nature* **411**, 421 (2001). doi:10.1038/35078161; Medline

31. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007). doi:10.1126/science.1136099; Medline

32. B. F. Jones, S. Wuchty, B. Uzzi, Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008). doi:10.1126/science.1158357; Medline

33. J. A. Evans, J. G. Foster, Metaknowledge. *Science* **331**, 721–725 (2011). doi:10.1126/science.1201765; Medline

34. J. Moody, D. R. White, Structural cohesion and embeddedness: A hierarchical concept of social groups. *Am. Soc. Rev.* **68**, 103–127 (2003).

35. R. D. Malmgren, J. M. Ottino, L. A. N. Amaral, The role of mentorship in protégé performance. *Nature* **465**, 622–626 (2010). doi:10.1038/nature09040; Medline

36. C.-T. Zhang, A proposal for calculating weighted citations based on author rank. *EMBO Rep.* **10**, 416–417 (2009). doi:10.1038/embor.2009.74

37. K. Börner, J. T. Maru, R. L. Goldstone, The simultaneous evolution of author and paper networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5266 (2004). doi:10.1073/pnas.0307625100; Medline

38. P. Deville, D. Wang, R. Sinatra, C. Song, V. D. Blondel, A.-L. Barabasi, Career on the move: Geography, stratification, and scientific impact. *Sci. Rep.* **4**, 4770 (2014). doi:10.1038/srep04770

39. A. Clauset, S. Arbesman, D. B. Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Sci. Adv.* **1**, e1400005 (2015). doi:10.1126/sciadv.1400005; Medline

40. J. Kaur, F. Radicchi, F. Menczer, Universality of scholarly impact metrics. *J. Informet.* **7**, 924–932 (2013). doi:10.1016/j.joi.2013.09.002

41. V. Larivière, C. Ni, Y. Gingras, B. Cronin, C. R. Sugimoto, Bibliometrics: Global gender disparities in science. *Nature* **504**, 211–213 (2013). doi:10.1038/504211a; Medline

42. S. F. Way, D. B. Larremore, A. Clauset, *Proceedings of the 25th International Conference on World Wide Web*, Geneva, Switzerland, 11 to 15 April 2016.

43. K. K. Mane, K. Börner, Mapping topics and topic bursts in PNAS. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5287–5290 (2004). doi:10.1073/pnas.0307626100; Medline

44. R. Sinatra, P. Deville, M. Szell, D. Wang, A.-L. Barabási, A century of physics. *Nat. Phys.* **11**, 791–796 (2015). doi:10.1038/nphys3494

45. S. Redner, Citation statistics from 110 years of physical review. *Phys. Today* **58**, 49–54 (2005). doi:10.1063/1.1996475

46. https://publish.aps.org/datasets.

47. T. Martin, B. Ball, B. Karrer, M. E. J. Newman, Coauthorship and citation patterns in the Physical Review. *Phys. Rev. E* **88**, 012814 (2013). doi:10.1103/PhysRevE.88.012814; Medline

48. J. P. Drenth, Multiple authorship: The contribution of senior authors. *JAMA* **280**, 219–221 (1998). doi:10.1001/jama.280.3.219; Medline

49. B. Cronin, Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *J. Am. Soc. Inf. Sci. Technol.* **52**, 558–569 (2001). doi:10.1002/asi.1097

50. S. Milojević, Principles of scientific research team formation and evolution. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3984–3989 (2014). doi:10.1073/pnas.1309723111; Medline

51. M. Levin, S. Krawczyk, S. Bethard, D. Jurafsky, Citation-based bootstrapping for large-scale author disambiguation. *J. Am. Soc. Inf. Sci. Technol.* **63**, 1030–1047 (2012). doi:10.1002/asi.22621

52. J. C. Lin, Chinese names containing a non-Chinese given name. *Cataloging & Classification Quarterly* **9**, 69–81 (1988). doi:10.1300/J104v09n01_08

53. S.-f. Lau, V. Wang, Chinese personal names and titles. *Cataloging & Classification Quarterly* **13**, 45–65 (2010). doi:10.1300/J104v13n02_04

54. H. Han, H. Zha, C. L. Giles, Digital Libraries, 2005. JCDL'05, Proceedings of the 5th ACM/IEEE-CS Joint Conference on (IEEE, 2005), pp. 334–343.

55. http://en.wikipedia.org/wiki/List_of_common_Chinese_surnames.

56. http://en.wikipedia.org/wiki/List_of_Korean_family_names.

57. V. I. Levenshtein, Soviet physics doklady (1966), vol. 10, pp. 707–710.

58. A. M. Petersen, W.-S. Jung, J.-S. Yang, H. E. Stanley, Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18–23 (2011). doi:10.1073/pnas.1016733108; Medline

59. A. M. Petersen, F. Wang, H. E. Stanley, Methods for measuring the citations and productivity of scientists across time and discipline. *Phys. Rev. E* **81**, 036114 (2010). doi:10.1103/PhysRevE.81.036114; Medline

60. F. Radicchi, C. Castellano, Rescaling citations of publications in physics. *Phys. Rev. E* **83**, 046116 (2011). doi:10.1103/PhysRevE.83.046116; Medline

61. S. Redner, How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* **4**, 131–134 (1998). doi:10.1007/s100510050359

62. M. J. Stringer, M. Sales-Pardo, L. A. Nunes Amaral, Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *J. Am. Soc. Inf. Sci. Technol.* **61**, 1377–1385 (2010). doi:10.1002/asi.21335; Medline

63. F. Radicchi, C. Castellano, Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *J. Informetr.* **6**, 121–130 (2012). doi:10.1016/j.joi.2011.09.002

64. L. Bornmann, H.-D. Daniel, What do citation counts measure? A review of studies on citing behavior. *J. Doc.* **64**, 45–80 (2008). doi:10.1108/00220410810844150

65. B. M. Althouse, J. D.West, C. T. Bergstrom, T. Bergstrom, Differences in impact factor across fields and over time. *J. Am. Soc. Inf. Sci. Technol.* **60**, 27–34 (2009). doi:10.1002/asi.20936

66. A. Schubert, T. Braun, Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics* **9**, 281–291 (1986). doi:10.1007/BF02017249

67. G. E. Box, G. M. Jenkins, G. C. Reinsel, in *Time Series Analysis: Forecasting and Control* (Wiley, ed. 4, 2013), 784 pp.

68. G. M. Beard, in *Legal Responsibility in Old Age* (Russells, New York, 1874), pp. 5–42.

69. S. Cole, Age and scientific performance. *Am. J. Sociol.* **84**, 958–977 (1979). doi:10.1086/226868

70. W. Dennis, Age and productivity among scientists. *Science* **123**, 724–725 (1956). doi:10.1126/science.123.3200.724; Medline

71. W. Dennis, Creative productivity between the ages of 20 and 80 years. *J. Gerontol.* **21**, 1–8 (1966). doi:10.1093/geronj/21.1.1; Medline

72. H. C. Lehman, in *Age and Achievement* (Princeton University Press, 1953).

73. B. Jones, E. Reedy, B. A. Weinberg, "Age and scientific genius" (Technical Report, National Bureau of Economic Research, 2014).

74. B. F. Jones, Age and great invention. *Rev. Econ. Stat.* **92**, 1–14 (2010). doi:10.1162/rest.2009.11724

75. B. A. Weinberg, D. W. Galenson, "Creative careers: The life cycles of Nobel laureates in economics" (Technical Report, National Bureau of Economic Research, 2005).

76. C. W. Adams, The age at which scientists do their best work. *Isis* **36**, 166–169 (1946). doi:10.1086/347941; Medline

77. H. Zuckerman, in *Scientific Elite: Nobel Laureates in the United States* (Transaction Publishers, 1977).

78. D. J. de Solla Price, Networks of scientific papers. *Science* **149**, 510–515 (1965). doi:10.1126/science.149.3683.510; Medline

79. J. A. Stewart, The Poisson-lognormal model for bibliometric/scientometric distributions. *Inf. Process. Manag.* **30**, 239–251 (1994). doi:10.1016/0306-4573(94)90067-1; Medline

80. N. L. Johnson, S. Kotz, N. Balakrishnan, in *Continuous Univariate Distributions* (John Wiley, vol. 1, 1994).

81. A. Clauset, C. R. Shalizi, M. E. Newman, Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009). doi:10.1137/070710111

82. E. J. Gumbel, in *Statistics of Extremes* (Dover Publications, 1958).

83. E. Limpert, W. A. Stahel, M. Abbt, Log-normal distributions across the sciences: Keys and clues. *Bioscience* **51**, 341–352 (2001). doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

84. M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions. *Internet Math.* **1**, 226–251 (2004). doi:10.1080/15427951.2004.10129088

85. N. L. Johnson, S. Kotz, N. Balakrishnan, Continuous multivariate distributions, in *Models and Applications* (John Wiley & Sons, vol. 1, 2002).

86. L. Egghe, The mathematical relation between the impact factor and the uncitedness factor. *Scientometrics* **76**, 117–123 (2008). doi:10.1007/s11192-007-1902-x

87. D. Stokols, K. L. Hall, B. K. Taylor, R. P. Moser, The science of team science. *Am. J. Prev. Med.* **35**, S77–S89 (2008). doi:10.1016/j.amepre.2008.05.002; Medline

88. K. Börner1, N. Contractor, H. J. Falk-Krzesinski, S. M. Fiore, K. L. Hall, J. Keyton, B. Spring, D. Stokols, W. Trochim, B. Uzzi, A multi-level systems perspective for the science of team science. *Sci. Transl. Med.* **2**, 49cm24 (2010). doi:10.1126/scitranslmed.3001399; Medline

89. T. Fawcett, An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006). doi:10.1016/j.patrec.2005.10.010

90. T. M. Cover, J. A. Thomas, in *Elements of Information Theory* (John Wiley & Sons, 2012).

91. E. L. Lehmann, G. Casella, in *Theory of Point Estimation* (Springer Science & Business Media, vol. 1, 1998).