Supplementary information to: Scientists have most impact when they're free to move

To accompany a Comment published in Nature ***550,*** *29–31 (2017);*
http://dx.doi.org/10.1038/550029a

Cassidy R. Sugimoto, Nicolas Robinson-Garcia, Dakota S. Murray, Alfredo Yegros-Yegros, Rodrigo Costas and Vincent Larivière

Supplementary Materials:

Data. We use the Web of Science database as our population frame. From these data, we extract a disambiguated dataset of individuals and their associated publications, curated using the disambiguation algorithm developed by Caron and van Eck (*1*). We restricted the analysis to the 2008 to 2015, given that Web of Science only provides full given names of authors and links to institutional affiliations from 2008. This makes author disambiguation more robust and allows for the linking of individual authors to locations. Our approach yields a dataset of 14,097,939 publications associated with 15,931,221 disambiguated individuals worldwide. For the analysis of papers' scholarly impact (citations), document type has been restricted to articles and reviews (a total set of 10,722,389 publications). Two subsamples are drawn from this set of researchers: those who have written their first paper between 2008-2012 (n=6,050,352 researchers), which allows to measure early career mobility, as well migrants and directional travelers who published their first paper 2008 onwards and published at least 8 papers until the end of 2015 (n= 12,046 researchers), for which mobility is tracked in a longitudinal manner. We consider all authors, regardless of author order.

Data validation. To analyze the validity of the approach for identifying international mobility, we compared our sample with the 2016 version of the Open Researchers and Contributor ID (ORCID) public data file, a public-access dataset where researchers can register for an identification number that uniquely associates them with their scientific contributions. This dataset has been used in previous studies to analyze scientific mobility (*3*). We first matched publications in our data with those in ORCID using unique identifiers (i.e., doi, pmid, WoS UTS). For those matched publications, we examine disambiguated authors from WoS with ORCID authors (based on surname, initial combinations, and e-mail addresses). When an ORCID author was linked to more than one disambiguated author in WoS, we assigned it based on (in order of priority): publications in common, number of publications, shared information in first name, recency of information, and breadth of information. If all things remain equal, a random assignment was made to a disambiguated author.

Using this matched analytic set, we identified only those ORCID IDs with demonstrated international mobility—that is, those who are affiliated with more than one country across both education and employment fields. ORCID profiles can be updated simultaneously with a mobility event, but bibliometric data takes longer to reflect changes: a researcher who migrates in a given year may not have a publication from that new affiliation for one or more years. Therefore, we match ORCID IDs from 2008-2013 with WoS data from 2008-2015. We filter for those individuals who have at least one publication in ORCID. From these, we find 9,522 matched researchers who have demonstrated international mobility in their ORCID profile. Of these, we find 63.6% to be mobile according to our bibliometric records. These rates fall within our expectation, given that the ORCID mobility indicators include movement across all ranks of education (bachelors, masters, professional, doctorate) as well as across non-scientific employment sectors. We would expect ORCID, therefore, to demonstrate higher degrees of mobility in comparison with a bibliometric approach, which captures only migration that is reflected in publications covered in the Web of Science.

Mobile scholars. Mobile scholars are defined as those who have had affiliations (on their Web of Science publications) in at least two distinct countries during the period studied. Non-mobile scholars are those who are affiliated to only one country over the period (within the WoS set of papers). Based on the entire set of scholars (n=15,931,221), 3.7% exhibit a form of mobility (n=595,894) and 96.3% are not mobile (n=15,335,327). Mobile scholars can be divided into three categories: migrants (n=162,519), directional travelers (n=213,810), and non-directional travelers (n=219,565). Migrants are defined as those scholars who move from one country to another, with at least a year where they are not affiliated to their country of origin. Directional travelers are those who, in addition to being affiliated to a new country, keep an affiliation to their country of origin. Non-directional travelers also have more than one country of affiliation throughout in the publication histories, but these co-affiliations appear consistently throughout the period, which makes it is impossible to determine a direction of travel. Both migrants and directional travelers can be considered as directional mobile scholars.

As our definition of mobility is based on affiliations found in papers, it is not independent from researchers' number of publications. Moreover, while migrants and directional travelers must produce at least two publications to be identified as such, non-mobility or non-directional mobility can be observed for scholars with a single publication. As expected, an increase in number of publications increases the likelihood of any form of mobility to be observed (Fig. S1). However, the proportion of various types of mobility varies according to the number of publications: while the proportion of migrants increases along with number of publications, it starts to decrease around 20 papers, mostly in favor of directional travelers, whose proportion increases steadily as a function of numbers of publications. The proportion of researchers classified as non-directional travelers increases until about 25 publications, and then begins to oscillate. On the whole, more than a third of scholars in our sample with 50 publications exhibit some form of mobility.

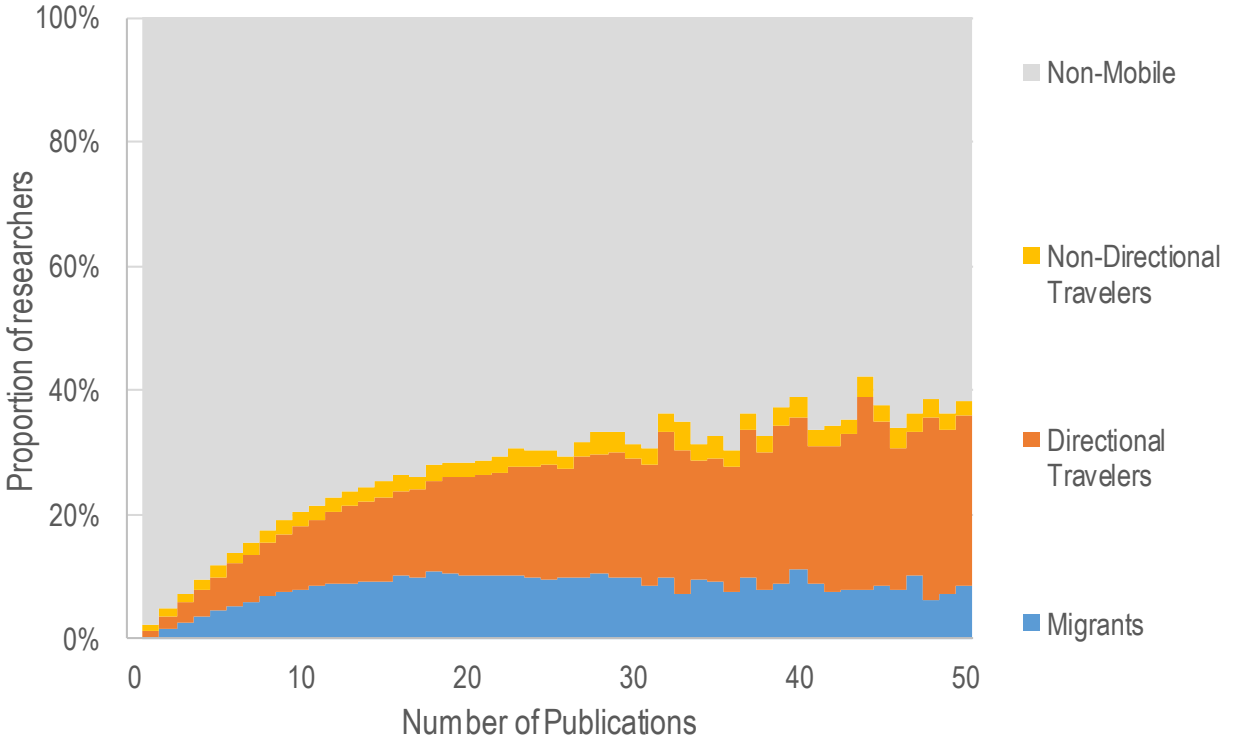

Fig. S1. Distribution of the proportion of researchers of each mobility type, by number of publications. For researchers who have written their first paper between 2008-2012. (n=6,050,352)

Mobility networks. International mobility networks are defined as countries that are linked by scholars who have been affiliated to them (not necessarily within the same publication, but during 2008-2015 period). The mobility network analysis is based on the global matrix of countries, linked by the number of scholars who have been affiliated to institutions from both countries. We count distinct scholars who have been affiliated to both countries at any point in time (here we consider all linkages between authors reported in publications through 2015). Fig S2 maps mobility networks for the complete set of 15,931,221 disambiguated author names. Countries that share many scholars will have strong linkages and therefore appear more strongly linked in the network analysis. The betweenness and closeness of centrality are provided by the top ten countries on each indicator (ranked by betweenness) for the world (Fig. S2A). Centrality measures are calculated for the complete set of researchers and countries using the open source software Gephi v.0.9.1. However, we used the free software VOSviewer v.1.6.5 (*4*) to visualize the network. Fig S2 only includes (for visualization purposes) those countries with at least 500 researchers. Maps use the VOS layout technique (*5*). Links between countries were normalized using the association strength method described in (*6*).

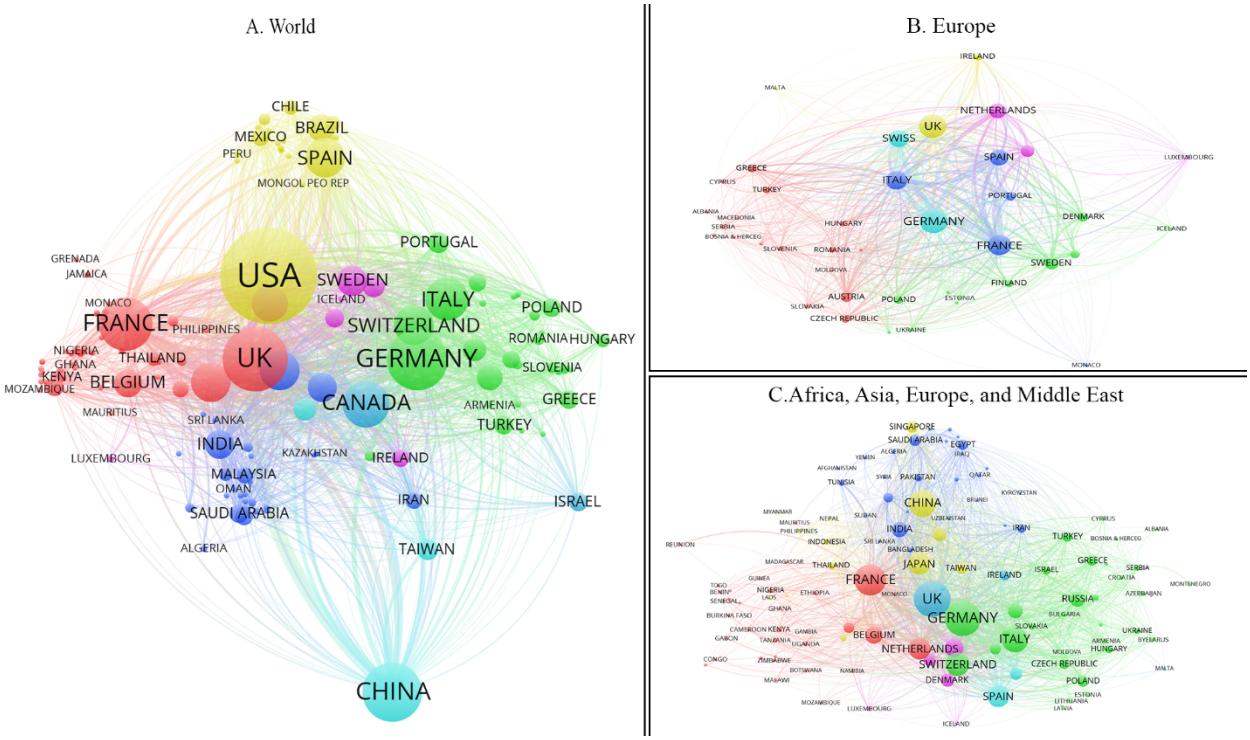

Fig. S2. Weighted mobility networks for countries with at least 500 researchers publishing between 2008-2015. Node size represents total number of links for a country. Link width represents the number of researchers affiliated with a pair of countries (either at the same time or sequentially). The colors denote community clusters (based on VOSviewer's clustering technique). (n=15,931,221)

Tab. S1. Betweenness and closeness of countries for the world map

Country	Closeness	Rank	Betweenness	Rank
USA	0.98484848	1	0.05158627	1
UNITED KINGDOM	0.94660194	2	0.03857314	2
FRANCE	0.9375	3	0.03718456	3
CANADA	0.91121495	4	0.03130858	4
GERMANY	0.91121495	5	0.02838484	6
SWITZERLAND	0.89041096	6	0.02389162	8
ITALY	0.88636364	7	0.03050356	5
NETHERLANDS	0.88636364	8	0.02430296	7
AUSTRALIA	0.87053571	9	0.0216997	9
BELGIUM	0.84782609	10	0.01914646	11
SPAIN	0.83333333	11	0.02077365	10

Tab. S2. Betweenness and closeness of countries for the European map

Country	Closeness	Rank	Betweenness	Rank
ITALY	0.977777778	1	0.034987741	1
FRANCE	0.977777778	2	0.0344592	2
SPAIN	0.956521739	3	0.031097175	3
TURKEY	0.936170213	7	0.028664045	4
GERMANY	0.956521739	4	0.011731927	5
SWITZERLAND	0.956521739	5	0.011731927	6
SWEDEN	0.956521739	6	0.011731927	7
NETHERLANDS	0.936170213	8	0.010479481	8
AUSTRIA	0.936170213	9	0.010450801	9
UNITED KINGDOM	0.936170213	10	0.009426985	10

Tab. S3. Betweenness and closeness of countries for the map of Africa, Asia, Europe, and the Middle East

Country	Closeness	Rank	Betweeness	Rank
FRANCE	0,954248366	1	0.045187238	1
ITALY	0,924050633	4	0.041075919	2
UNITED KINGDOM	0,948051948	2	0.038763404	3
GERMANY	0,941935484	3	0.033821573	4
SWITZERLAND	0,918238994	5	0.027352386	5
NETHERLANDS	0,906832298	6	0.026143859	6
SPAIN	0,834285714	13	0.023421763	7
PEOPLES R CHINA	0,85380117	10	0.019710065	8
BELGIUM	0,863905325	7	0.019156997	9
INDIA	0,863905325	8	0.018510044	10
SWEDEN	0,863905325	9	0.017936792	11

Analysis of flows between countries, regions, and continents. Figures "Making Tracks" and "Scientist Shuffle" (in the main manuscript) show flows of migrants and directional travelers who published their first paper in 2008 and published a total of at least 8 publications by the end of 2015 (n=12,046 researchers). Given that the number of researchers with co-affiliations increases over time, the number of researchers affiliated to each country grows over the 2008-2015 period. In order to correct for these issues, we fractionalized researchers by year and trajectory so that the total number of researchers during the period remains the same. Given that researchers do not have publications in each year, we take as their current affiliation the last known affiliation.

Tables S4-S6 present mobility flows of directional mobile scholars (migrants and directional travelers), by continent of origin and continent of destination, as well as between countries of the same continent. Table S5 presents the data as a function continent of origin, while S4 presents at a function of continent of destination. It shows that, while 23.9% of scholars have North America as their continent of origin, 35.6% of scholars end up having North America as their destination. On the other hand, 35.1% of scholars are originating from Europe, while 27.4% have this continent as their destination. A similar trend can be observed for Asia, where these percentages are of 27.1% and 21.8%, respectively. Africa and South America are the two continents that exhibit the highest proportion of within continent mobility. Typology of regions used in the paper is that of United Nations Statistics Division (https://unstats.un.org/unsd/methodology/m49/).

Tab. S4. Mobility flows between and within continents of directional mobile scholars with at least 8 publications (2008-2015) and whose first publication year was 2008. Number of researchers.

Continent of origin (2008)	Continent of destination							
	Africa	Asia	Europe	North America	Oceania	South America	All Continents	Diff. origin destination
Africa	98.1	73.2	127.6	72.2	5.4	1.4	377.9	12.8%
Asia	57.5	775.1	793.7	1433.4	183.4	24.6	3267.6	-19.5%
Europe	188.7	657.3	827.4	1993.6	327.0	229.5	4223.5	-21.7%
North America	70.8	1004.4	1082.6	424.8	169.5	131.5	2883.5	48.6%
Oceania	6.9	99.5	221.3	134.9	101.6	4.5	568.7	42.2%
South America	4.2	20.1	254.0	225.2	21.9	199.4	724.8	-18.5%
All Continents	426.2	2629.6	3306.6	4284.1	808.7	590.9	12046.0	-

Tab. S5. Mobility flows between and within continents of directional mobile scholars with at least 8 publications (2008-2015) and whose first publication year was 2008. Percentage calculated as a function of origin.

Continent of origin (2008)	Continent of destination						
	Africa	Asia	Europe	North America	Oceania	South America	All Continents
Africa	26.0%	19.4%	33.7%	19.1%	1.4%	0.4%	100.0%
Asia	1.8%	23.7%	24.3%	43.9%	5.6%	0.8%	100.0%
Europe	4.5%	15.6%	19.6%	47.2%	7.7%	5.4%	100.0%
North America	2.5%	34.8%	37.5%	14.7%	5.9%	4.6%	100.0%
Oceania	1.2%	17.5%	38.9%	23.7%	17.9%	0.8%	100.0%
South America	0.6%	2.8%	35.0%	31.1%	3.0%	27.5%	100.0%
All Continents	3.5%	21.8%	27.4%	35.6%	6.7%	4.9%	100.0%

Tab. S6. Mobility flows between and within continents of directional mobile scholars with at least 8 publications (2008-2015) and whose first publication year was 2008. Percentage calculated as a function of destination.

Continent of origin (2008)	Continent of destination						
	Africa	Asia	Europe	North America	Oceania	South America	All Continents
Africa	23.0%	2.8%	3.9%	1.7%	0.7%	0.2%	3.1%
Asia	13.5%	29.5%	24.0%	33.5%	22.7%	4.2%	27.1%
Europe	44.3%	25.0%	25.0%	46.5%	40.4%	38.8%	35.1%
North America	16.6%	38.2%	32.7%	9.9%	21.0%	22.3%	23.9%
Oceania	1.6%	3.8%	6.7%	3.1%	12.6%	0.8%	4.7%
South America	1.0%	0.8%	7.7%	5.3%	2.7%	33.7%	6.0%
All Continents	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Citation analysis. We calculate the mean normalized citation score (MNCS) of publications of individuals included in this study, considering their publications in the period 2008-2015 and counting citations (without self-citations) until week 29 of 2016, for the subset of scholars who have authored their first publication between 2008 and 2012 (n=6,050,352). Field-normalization at the paper level is determined by the journal subject categories of the publication journals of the researchers. As the unit analyzed here is researchers, MNCS at the researcher level are aggregated at the level of regions, which means that the results provided in Figure "Trip Adviser" (in the main manuscript) represent the average MNCS of each group of researchers rather than the average MNCS of the set of papers to which this set of authors contributed. For a broader discussion on the calculation of this indicator see Waltman et al. (*2*). Research production of scholars is obtained by averaging the mean number of articles and reviews to which researchers from each group contributed (full counting), irrespective of the authors' order. We also compiled the MNCS of the four categories of scholars as a function of their number of papers (Fig. S3). It shows that, for each level of research production, migrants have higher mean scholarly impact than travelers (directional and non-directional), and that non-mobile scholars obtain the lowest impact. On the whole, the mean impact of mobile scholars is 1.35, compared to

0.97 for non-migrants (39.6% advantage), which suggests that the greater the mobility, the higher the scholarly impact, and this advantage of is observed for every region.

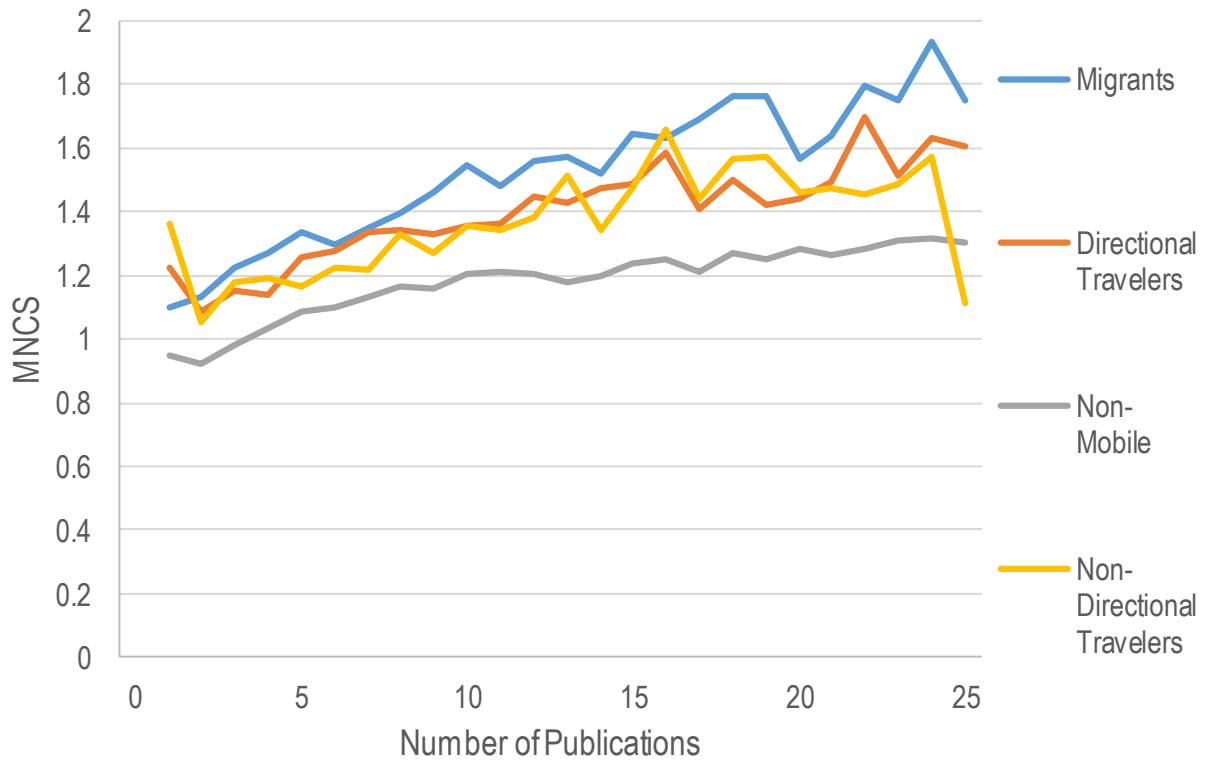

Fig. S3. Mean normalized citation score (MNCS) of each mobility type, by number of publications. For researchers who have written their first paper between 2008-2012.

Tab. S7. Number of researchers and MNCS of non-mobile and mobile researchers. For researchers who have written their first paper between 2008-2012.

Region	Non-mobile		Mobile		Gain in MNCS
	Researchers	MNCS	Researchers	MNCS	
Central and Western Asia	120,251	0.62	19,500	1.08	75.3%
Eastern Asia	2,194,819	0.85	67,485	1.30	52.7%
Eastern Europe	178,417	0.40	15,382	1.09	172.8%
Latin America and the Caribbean	266,330	0.54	25,729	1.01	87.8%
Northern Africa	41,335	0.56	7,254	0.84	50.8%
Northern America	1,285,555	1.43	128,996	1.59	10.8%
Northern Europe	348,735	1.19	76,300	1.52	27.1%
Oceania	120,012	1.11	23,453	1.44	29.4%
South-eastern Asia	86,427	1.01	18,564	1.23	22.6%
Southern Asia	271,921	0.56	25,779	1.01	78.3%
Southern Europe	332,164	0.77	48,064	1.32	72.0%
Sub-Saharan Africa	53,808	0.70	13,391	1.16	66.2%
Western Europe	536,300	1.03	119,219	1.52	47.0%
All regions*	5,836,074	0.97	214,278	1.35	39.6%

* The total number of mobile researchers is lower than the sum of scholars by region because of scholars affiliated with multiple regions

As a complement to Fig “Trip Adviser” in the main manuscript, we also compiled the MNCS pre- and post-mobility of mobile scholars, by region of origin and of destination (Fig S4A and S4B), as well as the difference in MNCS before and after the mobility event (Fig S4C) and associated numbers of researchers. This provides insights on the pairing of certain countries of origin and destination which yield the highest scholarly impact.

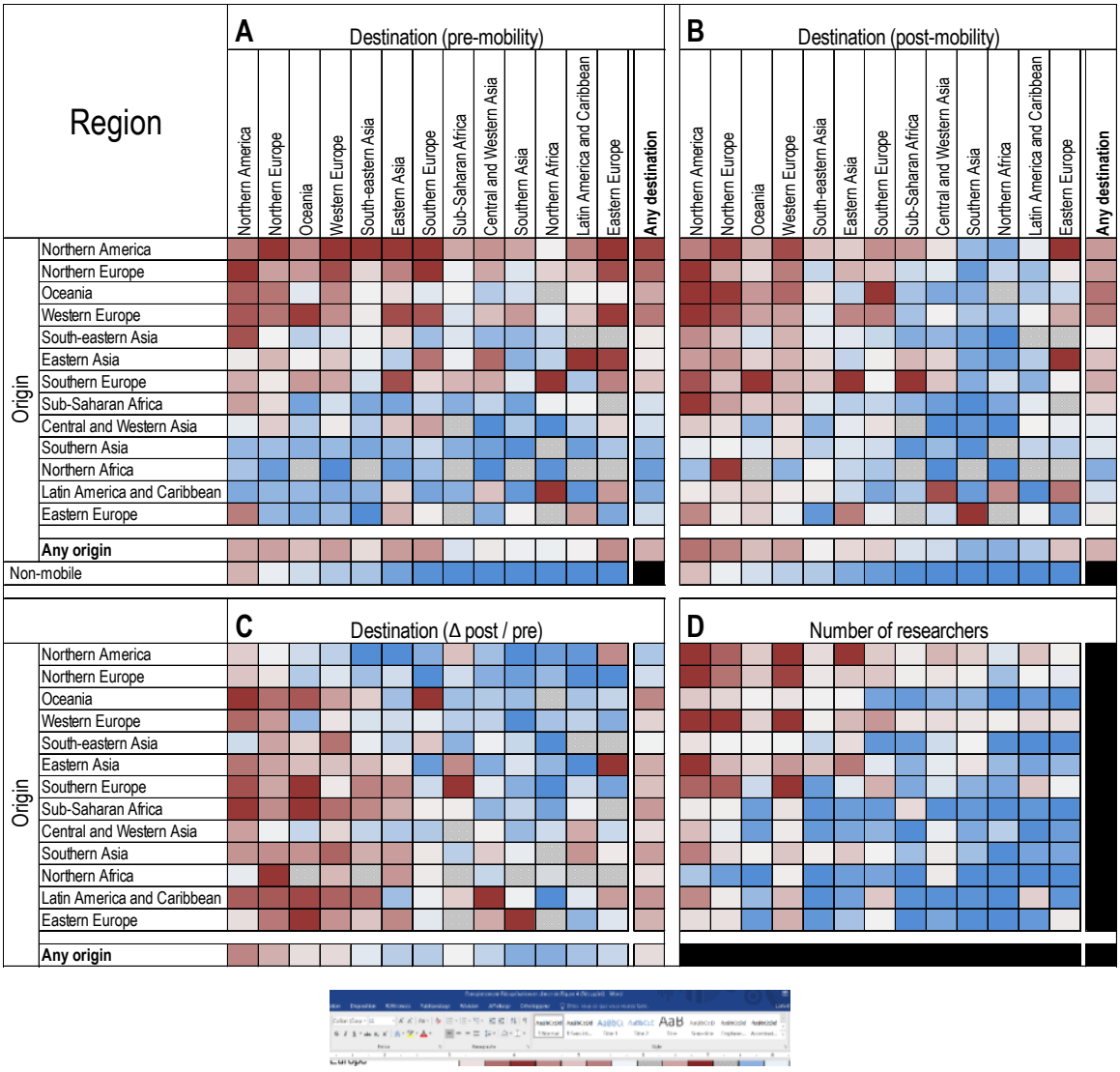

Fig. S4. MNCS of researchers, pre-mobility (A), post-mobility (B), difference between pre- and post-mobility MNCS (C), and number of researchers (D), by region of origin and of destination. Any origin provides the results for each region of destination, irrespective of origin, any destination provides the results for each region of origin, irrespective of destination. Diagonal denotes within region mobility. The bluest color represents the lowest frequency (fifth percentile); the 50th percentile is presented in white and the reddest are those within the top 5% of the distribution (95th percentile). Cells with gray lines have less than 30 mobile researchers.

Limitations of the suggested mobility approaches. Several limitations should be acknowledged in the analysis.

Coverage. The well-known limitations of bibliometric data apply here: Web of Science tends to be biased towards English-language material, in Western countries, in the natural and biomedical

sciences. Therefore, the mobility of some social sciences and humanities is underrepresented in these analyses. Furthermore, the use of journal articles alone is a limitation.

Granularity of time. Bibliometric data is updated quarterly, but the publication dates are only stable at the level of a year. Therefore, analyzing mobility monthly or weekly is not reliable using bibliometric data. Furthermore, given publication delays, the affiliation data on a given publication may not reflect the contemporary location of the scholar. Furthermore, publication delays vary by publication, so the ordering of the publication affiliations may distort reality.

First publication as proxy for origin country. For some of the analysis we restrict the data to those for whom the first publication was on or after 2008 and use the country/countries with which the scholars were affiliated as a proxy for origin. However, publication may begin when a scholar is abroad (e.g., for doctoral training), so these data should not be taken to imply that they are associated with the country in which the scholar was born (or funded by); rather, these data present the origin of the academic career of the scholar.

Author name disambiguation. The author-name disambiguation algorithm developed by Caron and van Eck (*1*) uses several variables to disambiguate authors, including the affiliation data of the authors. The algorithm lowers the probability that two publications belong to the same author when they are affiliated with different countries. Therefore, the publications of individuals with outlier patterns of behavior, such as rapid shifts across publication venues and disciplines, as well as set of co-authors may not be identified by the algorithm as belonging to a single individual. This means that our approach is actually a conservative one and likely underrepresents total worldwide mobility. Furthermore, although the disambiguation approach was published in a peer-reviewed conference proceeding, it has not been subsequently tested or validated, particularly in regards to non-English names.

Supplemental References

1. E. Caron, N. J. van Eck, Large scale author name disambiguation using rule-based scoring and clustering. Paper presented in the 19th International Conference on Science and Technology Indicators, Leiden, Netherlands, 3-5 September 2014.
2. L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, & A.F.J. van Raan, Towards a new crown indicators: an empirical analysis. Scientometrics, 87(3), 467-481, 2011.
3. J. Bohannon, Vast sets of Public CVs reveal the world's most migratory scientists. *Science News*. (May 2017).
4. http://www.vosviewer.com
5. N.J. van Eck, L. Waltman, Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538, 2010.
6. N.J. van Eck, L. Waltman, How to normalize co-occurrence data? An analysis of some well-known similarity measures. Journal of the Association for Information Science and Technology, 60(8), 1635-1651, 2009.