Supplementary Information

Foreign Born Scientists: Mobility Patterns for Sixteen Countries

1. Sampling procedure

We surveyed active researchers in the four scientific disciplines of biology, chemistry, earth and environmental sciences, and materials science during the period February-June 2011.

In order to construct the sample, we selected all journals classified by ISI as belonging to one of the four disciplinary fields and sorted them by Impact Factor for all subfields of the four disciplines. Impact Factor was taken from the latest available release of the Journal Citation Report of Thomson-Web of Science®. We then randomly picked a selection of four journals in each quartile of the Impact Factor distribution in each subfield of the four disciplines. In the aggregate, this selection corresponds to approximately 30% of all journals published in the four fields.

We downloaded full references of all scientific articles published in the selected journals in 2009 and retrieved the email address of the corresponding author. In case of multiple corresponding authors, we picked the first name on the list. In the case of corresponding authors appearing repeatedly in the list, we randomly selected one record. The four fields were chosen in part because 95 percent or more of all articles in these disciplines contain an email address for the corresponding author. More specifically, in 2009 the estimated number of records that did not report email address for corresponding author was 0.9% in biology, 3.6% in chemistry, 2.9% in earth and environmental sciences and 4.5% in materials science.

We coded the records by country, based on the domain of the email address (e.g. ".au" for Australia; ".be" for Belgium, etc.). We identified U.S. authors by those having ".edu" in the address, thereby restricting the U.S. sample to academic researchers.

Nature Biotechnology: doi:10.1038/nbt.2449

Surveyed countries are: Australia, Belgium, Brazil, Canada, Denmark, France, Germany, India, Italy, Japan, Netherlands, Spain, Sweden, Switzerland, United Kingdom, United States. China was initially included in the survey. However, a low response rate of less than 5 percent for a test sample of Chinese addresses suggested that respondents were either not receiving the invitation or had problems responding to the invitation. We thus decided not to survey researchers based in China.

The procedure produced a sample of unique article-corresponding author addresses of 47,304 scientists in 16 countries. The country panel sizes are highly variable. The smallest panel is Belgium (706) and the largest is the U.S. (14,059), reflecting the differentials in country contributions to scientific publications. Table S 1 provides summary statistics on the panel of invited respondents by country and scientific discipline.

Table S 1 Sixteen country panels by scientific discipline

COUNTRY	Biology	Chemistry	Earth	Materials Science	Total
Australia	470	386	490	225	**1571**
Belgium	253	214	131	108	**706**
Brazil	626	473	161	277	**1,537**
Canada	825	685	621	324	**2,455**
Denmark	189	170	99	55	**513**
France	1,026	1,380	671	762	**3,839**
Germany	1,303	1,533	763	781	**4,380**
India	282	587	160	351	**1,380**
Italy	771	1,097	514	397	**2,779**
Japan	1,485	1,996	562	1,207	**5,250**
Netherlands	382	275	223	156	**1,036**
Spain	620	939	369	375	**2,303**
Sweden	326	251	181	124	**882**
Switzerland	285	265	256	113	**919**
UK	1,312	1,051	748	584	**3,695**
U.S.	5,135	4,247	2,667	2,010	**14,059**
Total	**15,290**	**15,549**	**8,616**	**7,849**	**47,304**

Nature Biotechnology: doi:10.1038/nbt.2449

We track two characteristics of the articles from which the email of the corresponding author was extracted: the number of coauthors of the article and the total citations received by the article as of February 2010. The latter are retrieved from the Thomson-Web of Science® database.

Table S 2 Characteristics of the panels: total citations and number of co-authors

CORE COUNTRY	Total Citations*	Number of coauthors
Australia	0.95	4.61
Belgium	0.97	5.85
Brazil	0.39	5.27
Canada	0.85	4.33
Denmark	1.08	4.83
France	0.91	5.54
Germany	1.17	5.10
India	0.46	3.68
Italy	0.67	5.74
Japan	0.77	5.26
Netherlands	1.11	5.27
Spain	0.74	4.98
Sweden	1.01	4.88
Switzerland	1.55	5.14
UK	1.23	4.95
U.S.	1.31	4.64

*** Cumulated as of February 2010.**

Nature Biotechnology: doi:10.1038/nbt.2449

2. Survey administration and questionnaire

The main language of the survey is English. However, the questionnaire and the invitation emails were available in six other languages: French, German, Italian, Japanese, Portuguese and Spanish. The online questionnaire was developed through the platform Qualtrics® that supports multiple languages. The survey administrator chose a primary language to use in emails and set the list of languages available for a specific country survey. The platform then automatically deploys the language in which the recipient has set her browser, and lets the respondent switch from one language to another at any point while filling-in the questionnaire.

Table S 3 reports the country-languages used to administer the survey. Each panel member was emailed at most three times during February-June 2011 and asked to complete the web-based questionnaire. The platform recorded partial answers, allowing respondents to follow-up in additional rounds. The U.S. sample was divided into three blocks, due to the size of the sample. The questionnaire is available at http://www.dig.polimi.it/uploads/media/GlobSci_survey.pdf.

Table S 3 Primary and secondary languages in which the survey was administered

	Primary languages	Secondary languages
Australia	English	-
Belgium	English	-
Brazil	Brazilian Portuguese	English
Canada	English, French	-
Denmark	English	-
France	French	English
Germany	German	English
India	English	-
Italy	Italian	English
Japan	Japanese	English
Netherlands	English	-
Spain	Spanish	English
Sweden	English	-
Switzerland	English	French, German, Italian
UK	English	-
U.S.	English	-

Nature Biotechnology: doi:10.1038/nbt.2449

3. Response rates

Table S 4 reports the number of answers received by country. Answers are further divided into complete answers and partial answers. The latter are answers from respondents who began the survey, but dropped-out before reaching the last question. The total dropout rate is 5 percent. The response rate is 40.6 percent if both complete and partial answers are counted; 35.6 percent if only complete answers are counted. Reported response rates do not take into account undelivered invitations due to such things as incorrect email address, retirement or death and consequently underestimate the response rate.[1]

Table S 4 Response rates by country

	Panels	Total Answers	Of which complete	Of which dropout	Total Response Rate	Complete Response Rate
Australia	1,571	676	610	66	43.0%	38.8%
Belgium	706	302	244	58	42.8%	34.6%
Brazil	1,537	762	692	70	49.6%	45.0%
Canada	2,455	1,020	897	123	41.5%	36.5%
Denmark	513	227	208	19	44.2%	40.5%
France	3,839	1,618	1,367	251	42.1%	35.6%
Germany	4,380	1,326	1,147	179	30.3%	26.2%
India	1,380	627	484	143	45.4%	35.1%
Italy	2,779	1,917	1,759	158	69.0%	63.3%
Japan	5,250	1,860	1,678	182	35.4%	32.0%
Netherlands	1,036	391	345	46	37.7%	33.3%
Spain	2,303	1,228	1,080	148	53.3%	46.9%
Sweden	882	353	301	52	40.0%	34.1%
Switzerland	919	356	320	36	38.7%	34.8%
UK	3,695	1,355	1,183	172	36.7%	32.0%
U.S.	14,059	5,165	4,512	653	36.7%	32.1%
Total	**47,304**	**19,183**	**16,827**	**2,356**	**40.6%**	**35.6%**

The first panel of the table in the text is based on the 17,182 responses from scientists studying or working in one of the sixteen core countries in 2011 for whom country of residence at 18 is also know.

[1] Walsh, Cohen and Cho (2007) find in a sample of U.S. scientists that undelivered emails accounted for approximately 3.2 percent. Sauermann and Roach (2011) find that undelivered emails accounted for 6.3 percent in a sample of junior U.S. scientists.

The second panel of the table in the text is based on the subset of 15,115 scientists who lived in a core country at age 18 and provided full and consistent information on international experience(s).

Response rates by scientific field are reported in Table S 5. Participation was highest for scientists in earth and environmental sciences and lowest for scientists in biology. Differences in country and discipline participation are likely to reflect in part the degree to which similar populations of scientists have been surveyed in the recent past by other, unrelated studies.[2]

Table S 5 Response rates by scientific field

DISCIPLINE	Panels	Total Answers	Of which complete	Of which dropouts	Total Response Rate	Complete Response Rate
Biology	**15,290**	5,810	5,097	713	38.0%	33.3%
Chemistry	**15,549**	6,324	5,524	800	40.7%	35.5%
Earth & Environment	**8,616**	3,956	3,532	424	45.9%	41.0%
Materials Science	**7,849**	3,093	2,674	419	39.4%	34.1%
Total	**47,304**	**19,183**	**16,827**	**2,356**	**40.6%**	**35.6%**

4. Non-response bias

We assess non-response bias along three dimensions. First, we compare early and late respondents; second, we compare respondents against non-respondents and third, we compare full-respondents against those who dropped-out. Comparison is done for two characteristics known for the entire panel and sample: total citations received by the underlying article and number of coauthors. Total citations are likely positively correlated with the eminence of the scientist and could potentially reflect differentials in the propensity to answer related to how busy the respondent is. Because the number of coauthors was a basis for a branching question in the survey, more coauthors meant that more questions were asked. Therefore, it is potentially associated with dropping out of the survey.

[2] Haeeussler (2011) and Sauermann and Roach (2011) provide two recent examples.

Tests for equality of means are performed for each pair of country samples.

Table S 6 Two-groups comparisons. T-Tests. Hypothesized difference (early respondents – late respondents)=0

		Total Citations	Number of authors
Australia	mean diff.	0.174	-0.393
	st.err.	0.200	0.287
Belgium	mean diff.	-0.084	0.575
	st.err.	0.329	0.450
Brazil	mean diff.	0.148	0.144
	st.err.	0.083	0.227
Canada	mean diff.	-0.208	-0.372
	st.err.	0.132	0.219
Denmark	mean diff.	0.192	-0.367
	st.err.	0.481	0.519
France	mean diff.	0.047	-0.167
	st.err.	0.133	0.216
Germany	mean diff.	-0.140	-0.042
	st.err.	0.221	0.238
India	mean diff.	-0.093	0.119
	st.err.	0.117	0.217
Italy	mean diff.	-0.049	-0.334
	st.err.	0.117	0.230
Japan	mean diff.	0.151	-0.060
	st.err.	0.157	0.241
Netherlands	mean diff.	0.045	0.250
	st.err.	0.254	0.364
Spain	mean diff.	-0.099	-0.040
	st.err.	0.138	0.203
Sweden	mean diff.	-0.123	-0.531
	st.err.	0.317	0.448
Switzerland	mean diff.	-0.297	-0.357
	st.err.	0.438	0.438
UK	mean diff.	0.165	0.173
	st.err.	0.182	0.235
U.S.	mean diff.	0.199	0.074
	st.err.	0.106	0.102

***p<0.05**

Results of group comparisons for early and late respondents are reported in Table S 6. Early-respondents are those who completed the survey during the first and second round and late-respondents are those who completed the survey during third round. The results show no significant difference at the 5% confidence level.

Nature Biotechnology: doi:10.1038/nbt.2449

Comparisons of statistics for non-respondents against respondents are reported in Table S 7. A higher propensity to answer from authors with better-cited papers is found for France, Italy, Spain and the U.S. Authors of papers with more co-authors are also more likely to have answered from Brazil, Germany, Italy and the U.S. Thus, although there is some response bias, it is not in the direction that one might hypothesize.

Table S 7 Two-groups comparisons. T-Tests. Hypothesized difference (non-respondent – respondent)=0

		Total Cites	Number of authors
Australia	mean diff.	-0.039	0.035
	st.err.	0.098	0.142
Belgium	mean diff.	-0.268	-0.274
	st.err.	0.162	0.222
Brazil	mean diff.	0.088	0.397
	st.err.	0.046	0.125*
Canada	mean diff.	0.009	0.160
	st.err.	0.063	0.105
Denmark	mean diff.	-0.002	-0.114
	st.err.	0.224	0.242
France	mean diff.	0.122	0.029
	st.err.	0.058*	0.094
Germany	mean diff.	0.158	0.205
	st.err.	0.092	0.099*
India	mean diff.	0.029	0.008
	st.err.	0.052	0.096
Italy	mean diff.	0.181	0.288
	st.err.	0.061*	0.12*
Japan	mean diff.	0.089	0.112
	st.err.	0.052	0.080
Netherlands	mean diff.	0.069	0.031
	st.err.	0.124	0.178
Spain	mean diff.	0.161	0.051
	st.err.	0.064*	0.095
Sweden	mean diff.	-0.040	0.089
	st.err.	0.133	0.188
Switzerland	mean diff.	0.212	0.206
	st.err.	0.200	0.200
UK	mean diff.	0.143	0.123
	st.err.	0.083	0.108
U.S.	mean diff.	0.354	0.146
	st.err.	0.052*	0.049*

***p<0.05**

Nature Biotechnology: doi:10.1038/nbt.2449

Results of test comparisons for full-respondents against partial respondents (dropouts) are reported in Table S 8. Results indicate that more cited authors from Belgium were more likely to dropout before completing the study. The opposite is true for more cited authors from India, who were more likely than less-cited to take the survey in full. Dutch authors with more coauthors are also more likely to have completed the survey in full.

Table S 8 Two-groups comparisons. T-Tests. Hypothesized difference (complete – dropout)=0

		Total Citations	Number of authors
Australia	mean diff.	-0.162	-0.637
	st.err.	0.224	0.371
Belgium	mean diff.	-0.962	-0.120
	st.err.	0.405*	0.463
Brazil	mean diff.	-0.065	-0.298
	st.err.	0.104	0.299
Canada	mean diff.	0.168	-0.257
	st.err.	0.150	0.242
Denmark	mean diff.	0.029	-0.293
	st.err.	0.670	0.650
France	mean diff.	0.192	0.278
	st.err.	0.122	0.197
Germany	mean diff.	-0.096	-0.387
	st.err.	0.207	0.236
India	mean diff.	0.196	0.064
	st.err.	0.084*	0.175
Italy	mean diff.	-0.069	-0.417
	st.err.	0.105	0.239
Japan	mean diff.	0.176	0.079
	st.err.	0.144	0.214
Netherlands	mean diff.	0.565	0.872
	st.err.	0.290	0.438*
Spain	mean diff.	0.111	-0.068
	st.err.	0.117	0.192
Sweden	mean diff.	0.401	-0.161
	st.err.	0.300	0.387
Switzerland	mean diff.	-0.832	-0.479
	st.err.	0.517	0.455
UK	mean diff.	0.015	-0.063
	st.err.	0.175	0.282
U.S.	mean diff.	-0.130	0.105
	st.err.	0.101	0.118

*** p<0.05**

Nature Biotechnology: doi:10.1038/nbt.2449

5. Reasons for Return

All respondents who indicated that they had studied or worked in a country different form their country of origin, were asked whether or not they had subsequently returned.

Those who had returned, were asked to rate the importance of the reasons behind their decision to return. The rating scale ranged from 1 to 5 with steps of .1, where 1 was "totally unimportant" and 5 was "extremely important." Average scores assigned by returnees to each of the motivations by country of origin (and return) are reported in Table S 9.

Table S 9 How important was each of the following reasons behind your decision to return to the country where you lived when you were 18. Average score

COUNTRY OF ORIGIN	excellence / prestige of the institution in my area of research	outstanding faculty, colleagues or research team	better research infrastructures and facilities	greater availability of research funds	better wage / monetary compensation	better fringe benefits (parental leaves, pension, insurance, ..)	better working conditions (vacations, hours of work, ..)	opportunity to work with a specific group of scholars or colleagues	better job opportunity or career prospects	better quality of life	personal or family reasons	visa or immigration reasons
Australia	3.1	3.1	2.8	2.6	2.9	2.7	2.8	2.9	3.2	4.1	4.3	1.9
Belgium	2.8	2.9	2.6	2.5	2.4	2.9	2.8	3.3	3.4	3.5	4.4	1.8
Brazil	3.1	3.0	2.4	2.6	2.7	2.6	2.6	2.9	3.3	3.3	4.1	1.9
Canada	3.3	3.3	2.9	2.9	2.8	3.2	2.9	2.9	3.4	3.9	4.1	1.6
Denmark	2.8	3.1	2.7	2.7	2.8	3.1	3.1	2.8	3.1	3.5	4.5	1.3
France	3.4	3.4	2.8	2.4	2.2	3.6	3.3	3.4	3.4	3.8	4.2	1.5
Germany	3.4	3.5	3.2	3.3	2.8	3.1	2.7	3.2	3.6	3.3	4.0	1.6
India	4.0	3.5	3.3	3.3	2.8	3.0	3.1	3.1	3.4	3.3	4.1	2.1
Italy	2.6	2.6	1.9	1.8	1.8	2.2	2.2	2.7	2.6	3.2	4.3	1.4
Japan	3.2	3.2	3.3	3.1	3.0	2.8	2.7	3.2	3.6	3.2	3.7	2.0
Netherlands	2.9	3.1	2.8	2.6	2.6	2.6	2.6	2.9	3.3	3.1	4.1	1.6
Spain	2.8	3.0	2.3	2.3	2.3	2.6	2.6	3.0	3.4	3.7	4.5	1.4
Sweden	2.8	2.8	2.7	2.9	2.4	3.5	3.2	2.7	3.0	3.7	4.5	1.5
Switzerland	3.6	3.6	3.7	3.7	3.6	3.1	2.8	3.3	3.6	3.8	4.3	1.5
UK	3.6	3.5	2.9	2.9	2.4	2.5	2.6	3.2	3.4	2.9	3.9	1.8
USA	3.6	3.6	3.5	3.4	3.1	2.7	2.4	3.1	3.8	2.9	3.9	1.8
AVERAGE	**3.3**	**3.3**	**2.9**	**2.8**	**2.6**	**2.9**	**2.7**	**3.1**	**3.4**	**3.4**	**4.1**	**1.7**

Nature Biotechnology: doi:10.1038/nbt.2449

Table S 10 reports the percent that rated as “important” or “very important” the same motivations. Regardless of country, the most likely reason scientists give for returning to their country of origin is for personal or family reasons.

Table S 10 How important was each of the following reasons behind your decision to return to the country where you lived when you were 18. Percent indicating “important” or “extremely important”

COUNTRY OF ORIGIN	excellence / prestige of the institution in my area of research	outstanding faculty, colleagues or research team	better research infrastructures and facilities	greater availability of research funds	better wage / monetary compensation	better fringe benefits (parental leaves, pension, insurance, ..)	better working conditions (vacations, hours of work, ..)	opportunity to work with a specific group of scholars or colleagues	better job opportunity or career prospects	better quality of life	personal or family reasons	visa or immigration reasons
	%	%	%	%	%	%	%	%	%	%	%	%
Australia	33.6	27.9	15.6	11.6	23.3	26.0	22.4	27.0	35.7	68.4	83.3	6.4
Belgium	20.4	20.4	12.6	9.6	7.3	29.0	17.9	41.6	47.5	45.0	86.0	7.3
Brazil	40.9	34.9	13.6	22.4	22.6	22.2	19.1	29.1	42.8	46.0	75.4	12.0
Canada	43.0	40.6	21.5	23.4	22.6	40.2	28.0	25.7	44.3	66.5	73.4	5.5
Denmark	22.7	35.0	22.2	22.7	19.5	39.3	43.3	20.1	38.9	44.9	91.5	3.5
France	41.6	41.0	17.2	8.4	8.0	53.6	38.1	45.1	44.9	60.8	80.4	5.0
Germany	48.1	49.8	36.1	39.8	19.3	34.0	19.1	36.3	49.6	38.2	72.2	7.6
India	67.2	49.9	35.6	39.5	19.6	24.8	27.8	32.0	39.1	34.9	70.8	14.3
Italy	20.7	22.0	5.2	5.5	4.8	11.4	8.5	27.0	23.8	37.6	82.2	1.9
Japan	34.9	31.6	38.8	27.8	24.8	16.7	14.7	35.1	52.1	32.5	59.2	9.7
Netherlands	22.7	28.3	19.0	15.6	17.8	17.7	16.6	27.2	45.7	37.4	75.1	7.2
Spain	22.5	31.0	7.2	7.5	8.5	16.8	14.9	31.3	45.2	57.8	87.9	2.9
Sweden	23.8	21.8	22.8	23.1	6.4	48.1	32.4	22.7	26.3	52.4	87.5	1.3
Switzerland	55.6	52.3	56.6	56.8	50.4	35.7	18.8	42.2	48.7	60.7	82.7	3.2
UK	49.7	43.8	19.8	18.8	8.5	11.5	14.8	36.2	43.0	19.6	65.8	11.2
USA	56.6	53.9	50.4	44.9	36.1	20.6	13.3	34.2	62.6	24.3	66.2	10.7
AVERAGE	**40.5**	**38.9**	**25.9**	**23.9**	**18.1**	**26.8**	**20.4**	**33.9**	**45.3**	**43.5**	**74.5**	**7.5**

Nature Biotechnology: doi:10.1038/nbt.2449

6. Geographic diversity of foreign born scientists

The concentration rate presented in column 4 in the paper and reported in column 2 of Table S 11 measures the percent of a country's foreign scientific population coming from the top four countries. In order to provide a global measure of diversity of foreign born researchers and PhD students in each of the core countries, we compute the Herfindahl–Hirschman concentration index (H) of the shares of immigrants by current country (column 3, Table S 11). The H index is a standard indicator used to measure market concentration.

For each core country i, the corresponding H index is computed as $H_i = \sum_{j=1}^{C} \left(s_j^2\right) * 100$

where C is the total number of source countries and $s_j = N_j / N_{TOT}$ is the share of immigrants in country i from country j (N_j) with respect to the total number of immigrants in country i (N_{TOT}). Higher values of the H index indicate higher geographic concentration of the immigrants, with an upper bound value of 100. The lower bound of the indicator is given by H=(1/C)*100. This would be the case for a core country that has the same number of immigrants from all the possible source countries (127 countries in our dataset with a corresponding lower bound of 0.79). By construction the indicator tends to give more weight to larger shares.

As expected, the data reported in column two and three show a significant correlation (Spearman's rank correlation coefficient r=0.967). According to the H index, Germany is the most diversified of the 16 core countries and Switzerland is the most concentrated. Canada and Australia share virtually the same four-country concentration rate but Canada appears to have a relatively more diversified pattern of immigrants when we extend the analysis to all the source countries. The U.S. has the median value of diversification according to both the four-country concentration rate and the H index.

Table S 11 – Indicators of geographical concentration of foreign born scientists current working or studying in one of the core countries

CURRENT COUNTRY OF WORK OR STUDY	Four country concentration rate	H index of concentration
Australia	43.6	7.7
Belgium	52.2	8.7
Brazil	54.0	8.7
Canada	43.3	6.3
Denmark	44.5	9.4
France	37.2	5.9
Germany	30.2	4.2
India	100	*
Italy	42.6	6.8
Japan	60.5	14.8
Netherlands	40.6	6.2
Spain	40.2	6.2
Sweden	34.7	5.0
Switzerland	59.4	16.3
UK	37.6	5.8
USA	42.9	6.5

*** Numbers are too small to provide meaningful statistics**

7. References

Haeussler, C. *Research Policy.* 40:105 (2011).

Hirschman, A. "The Paternity of an Index." The American Economic Review 54(5):761, (1964).

Sauermann, H, and M Roach. "Not all scientists pay to be scientists : Heterogeneous preferences for publishing in industrial research," Georgia Institue of Technology, working paper (2010).

Walsh, J, W Cohen, and C Cho. *Research Policy* 36:1184 (2007).

Nature Biotechnology: doi:10.1038/nbt.2449

8. Acknowledgements

This research could not have been accomplished without the help and support of a large number of people and organizations. The authors acknowledge support from Regione Piemonte for the GlobSci project and from the IPE Program, National Bureau of Economic Research. Stephan acknowledges support from the European Commission (FP7) Project "An Observatorium for Science in Society Based in Social Models - SISOB" Contract no. FP7 266588 and Collegio Carlo Alberto Project "Researcher Mobility and Scientific Performance." Individuals who were especially helpful in the design phase are Henry Sauermann, John Walsh, James Poterba, Scott Stern, Sadao Nagaoka. For translations and beta testing we are indebted to Isabel Bodas Freitas, Sandro Battisti, Kun Zhang, Baoli Li, Diego D'Adda, Cristina Rossi-Lamastra, Sotaro Shibayama, Ana Herrero Alcalde, Seunghae Grace O, Michelle Perello, Tanja Schroot, Leslie Baksmaty, Alessandro Spadoni and Cornelia Meissner. Alessandro Fornari and Ali Mohammadi provided excellent research assistantship on data mining and coding. All errors are those of the authors.